Psychometrika

CONTENTS

(cont.)

VOLUME TWENTY-SIX DECEMBER 1961 NUMBER 4

Psychometrika

CONTENTS (Cont.)

NOTICE OF ANNUAL MEETING

Attention of all members of the Psychometric Society is invited to the 1962 Annual Meeting which will be held in St. Louis, Missouri, on September 3, 4, and 5. The Chairman of the Program Committee is Dr. John E. Milholland, Department of Psychology, University of Michigan, Ann Arbor. Suggestions as to symposia, with topics and possible names of participants, should be sent to Dr. Milholland as soon as possible. Since the Society will meet with the American Psychological Association, rules for symposia, contributed papers, and participation will be the same as those announced for that organization in *The American Psychologist.*

PSYCHOMETRIC MONOGRAPH NO. 9

The Psychometric Monographs Committee announces the publication of Stake, R. E. *Learning Parameters, Aptitudes, and Achievements.* Psychometric Monograph No. 9. Richmond, Va.: William Byrd Press, 1961, 70 pp., $2.00. Orders should be sent to Dr. John W. French, Educational Testing Service, P. O. Box 586, Princeton, New Jersey.

Truman Lee Kelley

Truman Lee Kelley

Truman Lee Kelley was born in Whitehall, Muskegon County, Michigan on May 25, 1884. He attended the University of Illinois where he received the A.B. Degree in 1909, and the A.M. in 1911. His early training was in mathematics and following his graduation from the University of Illinois in 1909, he became an instructor in mathematics at the Georgia Institute of Technology. He returned to the University of Illinois as an assistant in psychology to complete his Master's degree.

Dr. Kelley taught mathematics at Fresno (California) High School and Junior College and was a consulting psychologist at the Culver Military Academy prior to receiving his Ph.D. from Columbia University in 1914. His doctoral dissertation, titled *Educational Guidance*, foreshadowed the pattern of interests which directed his professional activities throughout his career. In this study he illustrated the use of the relatively new procedures of multiple correlation and regression coefficients as instruments for the type of educational guidance which has come into extensive use only in the last decade.

After obtaining his doctoral degree he was an instructor in educational psychology at the University of Texas and Teachers College, Columbia University until 1920, when he joined the faculty of Stanford University. During this period he also served as a psychological consultant to the Committee on Classification of Personnel, United States Army, and to the Surgeon General's Office in World War I. The years at Stanford University from 1920 to 1931 were very productive for Dr. Kelley. In 1923 the Stanford achievement Test Battery, of which he was a joint author, was first published. In 1924 the publication of his book, *Statistical Method,* marked an important milestone in the application of rigorous statistical methodology to problems in psychology, education, and other social science fields. In this book, as in all of his professional writings, Professor Kelley showed a passion for basic understanding and precise presentation.

In 1927 a book which soon became a classic in the educational field was published. This was Dr. Kelley's *Interpretation of Educational Measurements.* The following year, in 1928, *Crossroads in the Mind of Man* presented the evolution of his thinking on the problems of educational guidance between 1914 and 1928. This book extended Charles Spearman's tetrad tests to include a pentad function. It also proposed a substitution of a theory of intellect involving a number of dimensions in place of Spearman's theory of general intelligence which was very popular at the time. This publication represented an important landmark in aptitude testing and in many ways marked the beginning of a new phase in statistical analysis which has come to be known as "factor analysis."

In 1929 a series of Professor Kelley's lectures was published under the title, *Scientific Method.* The later phase of his career, during the period he was professor of education at the Harvard Graduate School of Education from 1931 to his retirement in 1950, was devoted primarily to more intensive studies of factor analysis, educational measurement, and statistical theory. In 1934 *Tests and Measurements in the Social Sciences* was published with Professor Kelley as a co-author. His solution of the principal components problem in factor analysis was published in his book *Essential Traits of Mental Life* in 1935. In 1938 *The Kelley's Statistical Tables* were first published. These very useful and widely known tables were revised and extended in the new edition which appeared in 1948.

Dr. Kelley's last major publication, *Fundamentals of Statistics,* was published in 1947. This book again clearly demonstrated Professor Kelley's quality of insight and insistence on thoroughness of treatment of basic issues.

During World War II Professor Kelley served as a consultant to the Secretary of War. He also directed a project on the development of an Activity Preference Test for the National Defense Research Committee.

Both before and after his retirement Professor Kelley was active in a wide variety of professional organizations. He was president of the Psychometric Society in 1938–39 and also served as a vice-president of the American Statistical Association. He was president of the Educational Research Corporation from 1946 to 1948. In 1946 he was one of the founders of the American Institute for Research and served as a member of its Board of Directors for more than ten years, including a three-year term as Chairman of the Board.

While at the University of Illinois, Professor Kelley was a co-founder of the national honorary education society, Kappa Delta Pi.

During his long and productive carrer Dr. Kelley found time for both mental and physical types of recreation. Many faculty members and students remember his enthusiastic participation in volley ball, tennis, and golf. He was an excellent chess player and during most of his career, an avid bridge player. In 1956 he earned the title of Life Master and was named a life member of the American Contract Bridge League.

Professor Kelley made many important contributions to both statistical and psychometrical theory and practice. His early efforts in statistics were focused on multiple correlation methods. His iterative method and his facilitating tables for computing partial correlation coefficients and regression equations were important aids to statisticians in all fields in the early 1920's. Another important contribution at a somewhat later stage was his development of epsilon, an unbiased correlation ratio measure. His measures of dispersion based on percentile ranges have become standard methods where this type of coefficient is applicable.

His wisdom and insight in dealing with statistics are well illustrated

by the following quotation from the first chapter, "The Dignity of Data and the Background of Statistics," in his book, *Fundamentals of Statistics*, published in 1947. "... applied statistics does not rest upon pure logic, ... but it is because it does not that it concerns itself with phenomena, not noumena, and that it is adaptable to all the problems of life in their partly repetitive aspects, which are their chief aspects. All the fine-spun deductions of mathematical statistics are by some amount wide of the mark when applied to real phenomena. Logically this must be so. They are astray by an amount which judgment alone bears witness to. How wide and what of it are questions that are very difficult of quantitative answer. As these difficulties are analyzed they regularly run back to the question of the soundness of some judgment of sameness or of relevance."

In psychometrics his main contributions have been in the area of factor analysis where he was one of the first Americans to advance the basic work of Spearman. He was a major contributor to the concepts of "true scores" and reliability coefficients and demonstrated the essential importance of these for interpreting test scores and other semi-reliable measures.

In test development his work on development of item weights and item validities provided rational procedures to substitute for the many empirical methods which had evolved. Similarly, his concept of "ridge route" norms provided a simple stable solution to an important practical problem. His basic point of view with respect to psychometrics was stated in the introductory chapter in *Crossroads in the Mind of Man*, published in 1928:

"Thus, in the field of psychology, if a designation of some trait or capacity, as a category of mental life, is to be given serious consideration, it must be such as to reveal itself as a measurable difference in conduct, that is, as a measurable difference in the same individual at different times, or in different individuals at the same time. ... This demand that a concept be subjected to objective measurement before it is worthy of serious consideration as an independent category of mental life, though sweeping, is not too sweeping, if we ... (include all) ... objective measurements ... (which) ... are definable and verifiable."

American Institute for Research *John C. Flanagan*

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

THE NATURE OF THE DATA, OR HOW TO CHOOSE A CORRELATION COEFFICIENT*

JOHN B. CARROLL
HARVARD UNIVERSITY

Some of our more self-critical psychometric brethren have reminded us (e.g., Dunlap [8]) that the purpose for which our society was founded—"the development of psychology as a quantitative rational science"—implies that our primary attention should be focussed on the formulation and testing of mathematical models for behavior, not on the elaboration of statistical methodology. I would very much like to have devoted this address to a new mathematical formulation of some aspect of behavior, something that would clearly fall within the purview of the Psychometric Society. Unfortunately, I have found myself encumbered with a good deal of unfinished business, both in and out of psychometrics, and time for mathematically formulating behavior has not been plentiful. But after reflection I have decided that some of my unfinished business does indeed have a solid connection with the rationales by which we quantify behavior, and it is about this that I wish to speak.

I am concerned with one of the most frequently used (perhaps also one of the most frequently misused) tools of psychometricians, namely, the correlation coefficient. It might be thought that after more than a half century of almost constant use, the correlation coefficient would have been thoroughly investigated and understood. But as recently as the years 1957 to 1959, the *American Psychologist* carried a discussion of "the needless assumption of normality in Pearson's r" [1, 10, 18, 21, 22]. The most recent article in this series, by Binder [1], was an astute resolution of the issues which had been raised, but it left untouched a number of problems relating to both the "Pearsonian" product-moment correlation and some of its derivatives. Also, I find that the presentation of the use of correlation measures in most textbooks in psychological and educational statistics leaves much to be desired. Some important matters are usually neglected, and the role of "assumptions" is often discussed in a way that will mislead the student. It is no wonder that contemporary research studies utilizing correlation measures sometimes suffer from various inadequacies in the use and interpretation of those measures.

*Presidential address delivered to the Psychometric Society, New York City, September 6, 1961.

It is also true that several otherwise highly distinguished textbooks in factor analysis give no discussion whatsoever of the problems in choosing appropriate correlation measures for entering in a correlation matrix. The one textbook which comes closest to giving a satisfactory account of the matter is Cattell's [4], but that account now needs to be updated. Gourlay [12] correctly identified the statistical artifacts which had arisen in several factorial studies because of improper choice of statistical measures of correlation, but John and Burt [16] found themselves unwilling to accept completely Gourlay's conclusions, perhaps because the argument was not sufficiently detailed. More recently Dingman [7] reported an investigation which, he claimed, failed to support the view that unequal dichotomization of items tends to produce spurious "difficulty factors." But as I hope to demonstrate here, Dingman's investigation was not an adequate test of this "view," and in any case, like the Spearman-Brown theorem on reliability and test length, this is not a view which needs to be empirically tested in order to be confirmed, but a theorem which can be demonstrated by purely mathematical techniques and which will obtain in empirical data to the extent that the data conform to the assumptions made in the theorem.

Recently in the course of some of my own research, I found myself faced with a serious question concerning the choice of a method of analysis for some rather unusual data which had been presented to me by Schuell and Jenkins—namely, data on a series of tests applied to 157 aphasic patients (cf., Schuell and Jenkins [24]). I chose to employ the tetrachoric correlation coefficient throughout this analysis, but arriving at this choice caused me to ponder more deeply the implications in such a procedure.

It has also occurred to me that the kinds of discussions of psychometric scaling initiated by Stevens [27] and carried on by such textbook writers as Siegel [26] and Senders [25] have been concerned almost exclusively with the properties of one scale at a time. These writers have neglected the possible bearing of inter-scale regressions; Luce's [20] important discussion considers such scale relations, but chiefly in psychophysical contexts.

Thus we have an important set of issues which do not seem to have been resolved in psychometric and statistical theory; at least, currently available information about correlation measures has not been widely disseminated. I propose to discuss some of this information. I shall, incidentally, be concerned only with the descriptive use of the correlation coefficient, for I feel that certain matters need to be set straight in the descriptive realm before bothering about inferential problems. Thus, with reference to recent papers by Norris and Hjelm [23] and by Heath [15] on the empirical sampling distributions of correlations involving non-normal distributions, there is a prior question about whether Pearsonian r's should be used for such distributions in the first place.

It is commonly recognized that the correlational method serves two main

functions: (1) as a basis for prediction from one variable to another, or from a set of variables to one or more dependent variables, and (2) as a way of measuring something called "relationship" between variables. Historically and logically, the former of these has held priority and correlational measures have been derived within the context of regression analysis. This paper is not concerned, however, with prediction; rather, it is concerned with the measurement of relations between variables. Prediction is something you do *after* you have discovered relationships between variables. But the prior measurement of relationships is important not only for prediction but also in its own right as a step in the construction of a science of behavior. Factor analytic theory, in stating that the correlation between two variables is the inner product of the respective vectors of factor coefficients, gives the correlation coefficient a meaning which is over and above that yielded by ordinary regression theory.

It is perhaps unfortunate that beginning students in psychology are usually first introduced to correlation theory via the well-known Pearsonian product-moment coefficient, which, they are told, characterizes the degree of relation between two variables and ranges in value between plus and minus one with intermediate values having various meanings which the instructor valiantly attempts to explain with such adjectival phrases as "moderate positive relationship," "substantial relationship," and "very high negative relationship." Students are not adequately informed that these limits and meanings strictly have reference to certain statistical models. Two of the most frequently used models are the normal bivariate surface and the linear regression model, but still other models are possible, as Binder [1] pointed out. No assumptions are necessary for the computation of a Pearsonian coefficient, but the interpretation of its meaning certainly depends upon the extent to which the data conform to an appropriate statistical model for making this interpretation. As the actual data depart from a fit to such a model, the limits of the correlation coefficient may contract, and the adjectival interpretations are less meaningful. The limiting case is provided when the two distributions are dichotomous and the points of dichotomy are asymmetrical between the two distributions, for here the Pearsonian coefficient (in this case, called the phi coefficient) does not, in general, range between plus and minus one, as Ferguson [9] showed some years ago. But even when the distributions have more than two class intervals, the possible range of the correlation coefficient is constricted to the extent that the two marginal distributions are disparate, i.e., not of identical shape and skew. The exact limits of the correlation coefficient in such a case can in fact be computed by techniques which I set forth in the course of a paper published in 1945 [3]; in that paper, I applied them to the special case of two tests with specified sets of items, but they are perfectly general, as shown in Appendix *A* here. In fact, they also give the limits for the point biserial *r*

range contracts

and the phi coefficient, although the limits for the phi coefficient are more conveniently obtained by using (for the positive case or for the reflected negative case) the formula

$$\phi_{max} = \sqrt{p_s q_l / p_l q_s}\ ,$$

where p_s and p_l are the respective proportions of successes, $p_s < p_l$, $p_s + q_s = p_l + q_l = 1$.

Figure 1, for example, depicts the highest degree of positive relationship which can be obtained between a certain pair of variables which are skewed highly in opposite directions; the Pearsonian correlation is .6037. (Table 1 shows the numerical data.) Therefore, the Pearsonian correlation computed between any two variables with this degree of disparity will be depressed in some degree by the fact of disparity.

The normal bivariate surface and the linear regression model have in common that the squares of the correlation coefficients yielded by them are both equal to the square of the correlation ratio, that is, to the complement of the ratio of the variances of errors of estimate to the total variance of the dependent variable. This suggests that the correlation ratio provides a more generally useful metric for interpreting degree of relationship than the correlation coefficient itself, since it provides a meaningful measure of relationship even when the data are of such a nature that a significantly

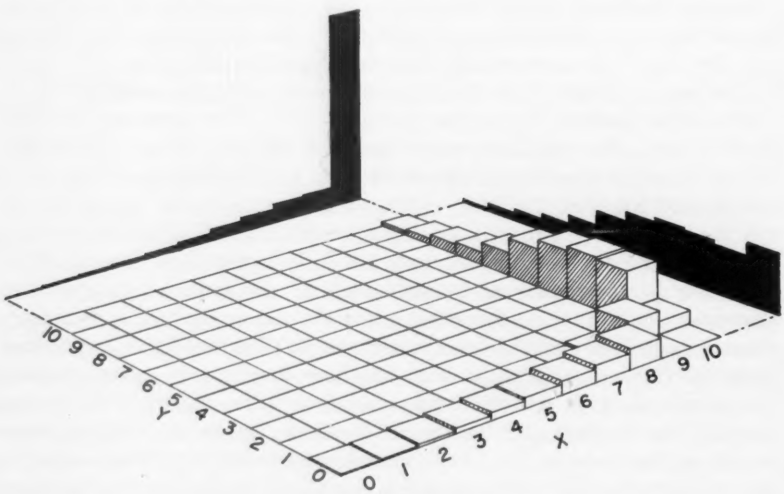

FIGURE 1

Bar Graph Showing the Maximum Possible Positive Relationship between Variables X and Y, Which are Skewed in Opposite Directions

TABLE 1
Joint Frequency Distribution for the Maximum Possible Positive Relationship between Variables X and Y, Which are Skewed in Opposite Directions (As Depicted in Figure 1)

Y	X 0	1	2	3	4	5	6	7	8	9	10	Sum
10	—	—	—	—	—	—	—	—	—	—	.0070	.0070
9	—	—	—	—	—	—	—	—	—	—	.0111	.0111
8	—	—	—	—	—	—	—	—	—	—	.0235	.0235
7	—	—	—	—	—	—	—	—	—	—	.0438	.0438
6	—	—	—	—	—	—	—	—	—	—	.0717	.0717
5	—	—	—	—	—	—	—	—	—	—	.1029	.1029
4	—	—	—	—	—	—	—	—	—	—	.1295	.1295
3	—	—	—	—	—	—	—	—	—	—	.1434	.1434
2	—	—	—	—	—	—	—	—	—	—	.1391	.1391
1	—	—	—	—	—	—	—	—	.0028	.0745	.0413	.1186
0	.0028	.0031	.0055	.0098	.0161	.0248	.0363	.0496	.0614	—	—	.2094
Sum	.0028	.0031	.0055	.0098	.0161	.0248	.0363	.0496	.0642	.0745	.7133	1.0000

Pearsonian $r = .6037$
Tetrachoric $r = 1.0000$
(for any cuts)

better fit can be attained with a nonlinear regression line. On the other hand, Guttman [14] has pointed out that factor analysis is stochastically justified only if the regressions are linear. Transformations of the marginal distributions in order to utilize the linear regression model may therefore be desirable. If the regression lines are both monotonic, it is always possible to make monotonic transformations of one or both of the variables to produce a linear regression line; even if a non-monotonic regression line is found, it may be desirable to make a non-monotonic transformation of a marginal distribution. In any case, use of the linear regression model implies that we are measuring degree of relationship with a coefficient which has the maximum and minimum values, and the interpretations, provided by the Pearsonian correlation coefficient.

Whatever we do with data by way of transforming distributions, however, we are still dealing with what may be called *manifest* relationships, that is, relationships which can be directly computed from the data. In building a quantitative rational science of psychology, we would be interested not so much in these manifest relationships but rather in the latent relationships which may be inferred to exist between variables and which are masked or distorted by various kinds of errors and constraints. To find a simple ex-

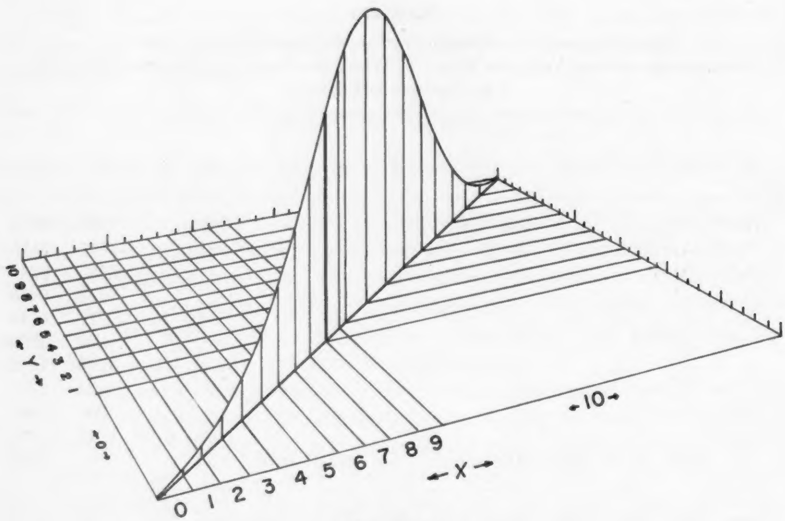

FIGURE 2
Diagram Showing How the Frequency Distributions of Figure 1 Can Arise from an Underlying Perfect Linear Relationship (r = 1.00)

ample, let us again examine Figure 1, where it may be supposed that errors of scaling and censoring of information have concealed the true relationship depicted in Figure 2. The true relation appears to be expressed by a straight line, and if we appeal to the model of linear regression theory we can express it by a coefficient of $+1$.

A study of the ways in which various types of errors and constraints affect assumed models of correlation surfaces and yield systematically biased joint distributions of manifest data will aid us in making better estimates of latent relationships. It should be noted that the relation between manifest and latent relationships between variables is not analogous to the relation between sample statistic and population parameter. Indeed, the distinction between manifest and latent relationships holds both for sample space and population space. In the discussion that follows, we can be referring to either sample space or population space, although it may be understood that problems of estimation are more severe in the former.

It is useful to distinguish between three general types of constraints which distort latent relationships.

1. *Errors of scaling*

On the assumption that there does indeed exist a "true" scale of at least an equal-interval character, we may identify errors which cause the

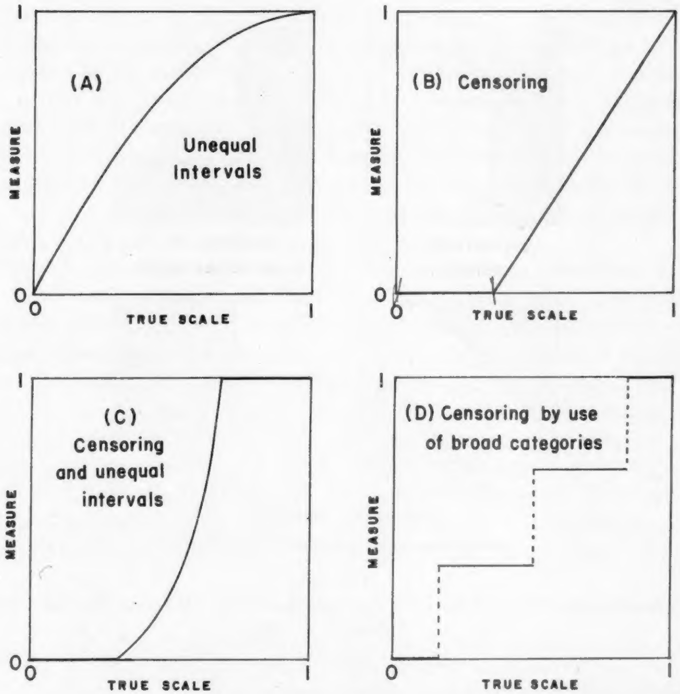

FIGURE 3
Examples of Several Types of Scaling Errors

scaling of manifest data to be a nonlinear transformation of the true scale. Errors of scaling thus include the error known among statisticians as censoring, that is, reporting all values within a certain range on the true scale as if they were all equal to the same value. Grouping of values in broad categories can be regarded as a variety of censoring. Figure 3 shows a number of types of scaling error; in each graph, the relation between the true scale and the manifest scale is nonlinear. We have restricted consideration to monotonic cases.

2. *Errors of scale-dependent selection*

Such errors occur whenever, through either explicit or implicit selection, certain portions of the frequency distribution of a variable are absent from the sample or from a statistical population, the portions not being independent of their locations on the scales. One variety of selection error is truncation, illustrated for the X variable in Figure 4A.

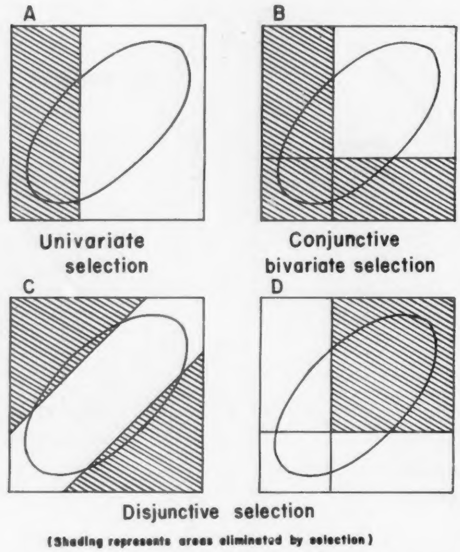

FIGURE 4
Examples of Several Types of Selective Processes Applied to a Bivariate Normal Correlation Surface ($r = .60$)

3. *Errors of measurement*

I have found it useful to distinguish between two types of measurement error: (a) *scedastic* error—random, unsystematic error arising from various sources such as sampling errors in the choice of test items, variation in psychophysical response criteria, etc.; and (b) *topastic* error—the partly random, partly systematic error which results when the individual taking a multiple-choice psychological test has the opportunity to get some of his answers correct by guessing.*

Let us review briefly the effects of these errors on joint distributions of variables and hence upon measures of correlation. Errors of scaling will always depress linear relationships, if present, and will at least disturb non-linear relationships.

Errors of selection may either increase or decrease measured relationships—usually the latter. Conjunctive bivariate selection, represented in Figure 4B, will always decrease the measured relationship. Figures 4C and 4D represent two types of what may be called *disjunctive* bivariate selection, one increasing the size of the relationship, the other markedly decreasing it.

*The word *scedastic* is derived from the Greek stem for "scattered"; *topastic* has been taken from the Greek root for "aimed at, guessed at."

The type represented in Figure 4C can occur "in real life" whenever cases that are imbalanced in their profile of characteristics have a lowered probability of appearing in the sample to be studied; for example, if people who are too skinny or too fat are excluded from a sample, the correlation between height and weight will be increased. The second type of disjunctive bivariate selection can easily occur in studies of symptomatology: for example, if we selected a large sample of individuals who are either blind *or* deaf or both, we would find a negative correlation between blindness and deafness even though it is reasonable to suppose them to be independent.

Scedastic errors of measurement affect correlations according to the familiar theorems of true score and error which have been developed in mental test theory, and the magnitude of the effects can be estimated by Spearman's formula for the correction for attenuation. Formulas for attenuation correction can also be applied to *topastic* errors of measurement, but, as I showed in 1945 [3], these errors also have systematic effects on marginal distributions which in turn can give spurious measures of correlation even when the tetrachoric correlation coefficient is used. Tables 2 and 3 illustrate the nature of these effects. The "true" or "latent" relationship between two tests is assumed to be that shown in Figure 2. Then the relationship would be as shown in Table 1 if there were neither scedastic nor topastic error; it would

TABLE 2

Joint Frequency Distribution of Variables X and Y of Figure 1 (and Table 1) but with Scedastic Error Superimposed

						X						
Y	0	1	2	3	4	5	6	7	8	9	10	Sum
10	—	—	—	—	—	—	—	—	—	—	.0151	.0151
9	—	—	—	—	—	—	—	—	—	—	.0208	.0208
8	—	—	—	—	—	—	—	—	—	—	.0344	.0344
7	—	—	—	—	—	—	—	—	—	—	.0516	.0516
6	—	—	—	—	—	—	—	—	—	—	.0705	.0705
5	—	—	—	—	—	—	—	—	—	.0001	.0891	.0892
4	—	—	—	—	—	—	—	.0001	.0002	.0006	.1043	.1052
3	—	—	—	—	—	—	.0001	.0004	.0011	.0024	.1119	.1159
2	—	—	—	—	—	.0001	.0004	.0012	.0034	.0074	.1067	.1192
1	—	—	—	—	.0001	.0004	.0013	.0037	.0086	.0159	.0846	.1146
0	.0048	.0050	.0081	.0124	.0178	.0243	.0310	.0365	.0381	.0349	.0505	.2634
Sum	.0048	.0050	.0081	.0124	.0179	.0248	.0328	.0419	.0514	.0613	.7395	.9999

Pearsonian r = .4974
Tetrachoric r = .92
(cuts nearest medians)

TABLE 3

Joint Frequency Distribution of Variables X and Y of Figure 1 (and Table 1) but with both Scedastic and Topastic Error Superimposed

					X							
Y	0	1	2	3	4	5	6	7	8	9	10	Sum
10	—	—	—	—	—	—	—	—	.0001	.0001	.0510	.0512
9	—	—	—	—	—	.0001	.0002	.0003	.0006	.0012	.1005	.1029
8	—	—	—	.0001	.0002	.0004	.0008	.0015	.0028	.0049	.1480	.1587
7	—	—	—	.0002	.0004	.0010	.0021	.0040	.0071	.0118	.1657	.1923
6	—	—	.0001	.0003	.0007	.0018	.0037	.0069	.0119	.0189	.1453	.1896
5	—	—	.0001	.0003	.0009	.0021	.0044	.0081	.0138	.0210	.1000	.1507
4	—	—	.0001	.0003	.0007	.0018	.0037	.0067	.0112	.0163	.0531	.0939
3	—	—	—	.0002	.0004	.0010	.0021	.0038	.0062	.0087	.0210	.0434
2	—	—	—	.0001	.0002	.0004	.0008	.0014	.0023	.0031	.0059	.0142
1	—	—	—	—	—	.0001	.0002	.0003	.0005	.0007	.0010	.0028
0	—	—	—	—	—	—	—	—	.0001	.0001	.0001	.0003
Sum	—	—	.0003	.0015	.0035	.0087	.0180	.0330	.0566	.0868	.7916	1.0000

Pearsonian $r = .3160$
Tetrachoric $r = .56$
(cuts nearest medians)

be approximately as shown in Table 2 if there were scedastic error alone, and it would be approximately as shown in Table 3 if both scedastic and topastic error are present. (Table 3 has been constructed for the case where c, the probability that an individual who "does not know" an answer will nevertheless select it, is $c = .5$.) Pearsonian correlations for all these cases are shown in the tables; also, tetrachoric correlations. It should be remembered that the assumed latent relationship between X and Y is represented by a coefficient of unity.

The manner in which these various constraints affect correlation coefficients becomes of critical importance in factor analysis. I understand factor analysis to be a technique for analyzing underlying dimensions of behavior; it should hence be concerned with what I call latent relationships. There is no particular point in making a factor analysis of a matrix of raw correlation coefficients when these coefficients represent manifest relationships which mask and distort latent relationships. Yet, there are numerous factor analytic studies which have tried to investigate such matrices.

Strictly speaking, this remark may be taken to apply to all factor analytic studies in which corrections for attenuation have not been made, and this includes practically every study that has ever been reported. There are, of course, two good reasons why attenuation corrections are not commonly

made in factor analytic studies: (a) as is well known, corrections for attenuation would not change the rank of the correlation matrix and would produce only proportional changes in the results, and (b) errors in estimating reliability coefficients would introduce further error into the data.

If all the effects of constraints on correlation coefficients were merely on the order of unreliability effects, we could ignore them. But they are not. It has been demonstrated that many of these constraints operate to alter the rank of a correlation matrix in important ways. Ferguson [9] was evidently the first to discover that the use of phi coefficients with a set of items heterogeneous in "difficulty" but containing a single "content" factor would preclude the finding of rank one in the correlation matrix (except under severely circumscribed conditions). It can be deduced from the demonstrations in [3] that this is also true whenever Pearsonian correlations are based on disparate marginal distributions. Except for the effects of purely random errors of measurement, it appears that all the types of constraints listed above can disturb the rank of a correlation matrix. If we are going to continue to make factor analytic studies, we should attempt to correct or adjust for these effects.

Granted that we wish to make these adjustments, how shall we proceed? The desirability of such adjustments was recognized fairly early in the history of factor analysis. In his first large study of primary mental abilities, Thurstone utilized tetrachoric correlations, partly in order to reduce computational labor, but partly because "the simplest psychological assumption that can be made is that each of the primary abilities is distributed normally in the experimental population. . . . Any linear function of the standard scores [hence, the test scores] will also be normally distributed . . . in using tetrachoric coefficients, we are estimating the product-moment coefficient for the normalized distributions of scores" ([29], pp. 58–59). My own factor-analytic study of verbal abilities was the first, to my knowledge, to use what was essentially a stanine normalization of raw scores, but though I still believe this was a defensible procedure I must plead guilty to having made the possibly misleading statement that "the assumptions underlying the product-moment correlation coefficient justify this step" ([2], p. 289). ?

It seems reasonable to assert that the kinds of adjustments that are to be made to estimate degree of latent relationship must be selected in the light of the investigator's analysis of what kinds of errors are to be adjusted for. The nature of the data must be scrutinized, and thought about, in order to glean evidence concerning rank-distorting constraints. It may be assumed that some random error of measurement (scedastic error) is present in all data, but since it is not rank-distorting it is convenient to ignore it.

Topastic error is a function of the nature of the measurement procedure; in its commonest form, it occurs when *S*s are allowed to choose their answers from a number of offered alternatives in such a way that there is a certain

probability that they will choose answers scored correct even when they do not know the correct answer. Topastic error is rank-distorting. The chief problem in correcting for it is that of estimating the value of c, the probability of chance success by guessing. In general, c cannot be taken as equal to the reciprocal of the total number of alternatives (a) in an item because of variation in the attractiveness of options. Scrutiny of certain empirical studies of comparable tests given under free-response and multiple-choice test conditions suggests that the value of c is to be estimated as somewhat *less* than the quantity $1/a$.* Apparently, *S*s who do not know the right answer are sufficiently well attracted by wrong answers to have less chance of guessing the right answer than a priori theory would allow. This conclusion may seem to contradict the advice often given to test-takers to "give your best guess," but it actually does not; it is a consequence of the fact that individuals who do not even have partial knowledge are less likely to select a correct answer than those who have partial knowledge. In any case, once the value of c (or the average value for a set of items) has been chosen, it is possible to adjust univariate and bivariate distributions for the effects of topastic error. The adjustment procedures for dichotomous items were given in [3], and the general procedures for tests with any number of items are given in Appendix *B* of this paper. It should be noted, incidentally, that these procedures have an effect which is radically different from that of the well-known procedure of correcting individual scores for chance success. The latter is a simple linear transformation of the raw scores and (unless there are an appreciable number of omits) will have little effect on correlation coefficients other than incidental effects of changing class intervals or points of dichotomization. The proper correction involves, in effect, estimating the probable univariate and bivariate distributions of non-topastically affected scores which would give rise to the observed distributions of topastically affected scores.

Thus, one error which Dingman [7] made in his attempt to investigate the relation between coefficients of correlation and difficulty factors was his failure to correct *distributions* of scores rather than merely *scores*.

Scedastic and topastic errors should therefore give us relatively little trouble. The real quagmire in this business is that of adjusting for errors of scale and errors of selection, because (a) these errors can have similar, and hence, indistinguishable effects in the observed data, and (b) the proper adjustments to make for these two effects may be precisely opposite to each

*Consider, for example, the data of Gage and Damrin [11], who gave parallel 45-item multiple-choice same-opposites tests with 2, 3, 4, and 5 choices to comparable groups. The mean *error* scores ($\bar{E}_c$) were 14.94, 20.45, 23.77, and 25.77, respectively. Extrapolating, let us assume that the mean error score would have been 27 if there had been an infinite number of choices, and this could be taken as equal to the "true," non-topastically affected value $\bar{E}$. Then, using formula (31) in [3], we find that the values of c for the tests with 2, 3, 4, and 5 choices would have been .445, .243, .120, and .046, respectively, as compared with the *a priori* values of .500, .333, .250, and .200, respectively. Further studies like Gage and Damrin's are needed to establish representative values of c for various types of items.

other. For example, single truncation (a variety of selection effect) may produce a distribution which suggests a scaling error, but trying to adjust for this effect by making a normalization transformation is making the situation worse rather than better. One can sink deep rather fast. The basic difficulty is that we are usually in ignorance as to the true nature of the underlying scales of our measurements, and we seldom have any accurate information as to how our samples are selected from the larger population. With regard to underlying scales of measurement, many believe that we have little justification for regarding them as more than ordinal in nature. Is there any help for us as we sink into this quagmire?

Errors of *conjunctive* selection have already been treated by Thurstone ([30], ch. XIX) and Thomson [28]. Errors arising from *disjunctive* selection would appear to have some odd effects, and to my knowledge they have never been investigated. In any case, without a knowledge of the manner in which a sample has been selected it is unlikely that one could correct for selection effects intelligently. In practice, one tries to obtain samples which are as representative as possible. On the other hand, there are situations in which one wishes to make studies of special groups. Consider the problem I faced when confronted with data on 157 aphasic patients. In the first place, aphasia is an extremely rare phenomenon in the total population. Any attempt to add normal subjects to the sample so that normals and aphasics would be represented in proportion to their incidence in the total population would be sheer madness. There was no alternative but to restrict attention to the sample of aphasic patients; after all, the problem was to determine the dimensionality of aphasic symptomatology. There was indeed the risk of obtaining correlations distorted by the effects of *disjunctive* selection (of the type shown in Figure 4D), but some consolation could be taken in the fact that any such effects would probably manifest themselves as *lowered* correlations between measures of *different* factors, without significantly disturbing correlations between measures of the same factor, and therefore the clarity of the factorial structure would be accentuated rather than blurred. One might even expect negative correlations between factors, and if such were obtained, they might be interpreted as an effect of disjunctive selection. As it turned out, negative correlations in the correlation matrix were extremely rare, and the correlations between the factors were generally positive. In any case, the data we had were such as to indicate that if there was any disjunctive selection, it had no seriously distorting effect on factorial structure.

Let us, therefore, turn our attention to scaling error. The commonest means of correcting for scaling error has been the transformation of distributions to an approximately normal form. This at least has the advantage that it converts distributions in such a way that they are approximately homogeneous in shape, thus precluding the errors in correlation matrices which arise from disparity of shape. It does not, of course, guarantee that the

Normal. Transformation

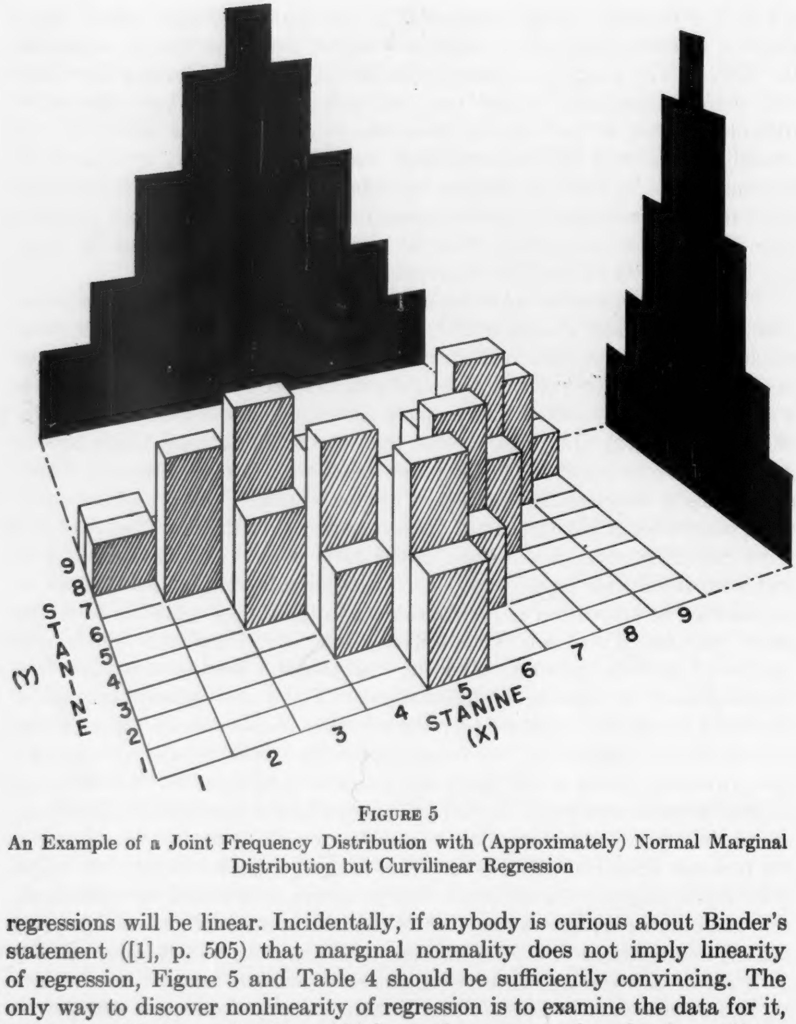

FIGURE 5

An Example of a Joint Frequency Distribution with (Approximately) Normal Marginal Distribution but Curvilinear Regression

regressions will be linear. Incidentally, if anybody is curious about Binder's statement ([1], p. 505) that marginal normality does not imply linearity of regression, Figure 5 and Table 4 should be sufficiently convincing. The only way to discover nonlinearity of regression is to examine the data for it, and even with the availability of high-speed computers there has been all too little examination of data in this respect. The explanation for this neglect may be found in the fact that those few who have taken the trouble to do so have not very often been rewarded by the discovery of significant nonlinearity (but see DeSoto and Kuethe [6]). One is tempted to conclude that the high incidence of linear regressions in psychological test data, at least, suggests that equal-interval scales are commoner than one might be led to

TABLE 4
A Joint Frequency Distribution with Approximately
Normal Marginal Distributions but Curvilinear Regression
(As Depicted in Figure 5)

	X									
Y	1	2	3	4	5	6	7	8	9	Sum
9	.020	—	—	—	—	—	—	—	.020	.040
8	.020	.015	—	—	—	—	—	.015	.020	.070
7	—	.055	.005	—	—	—	.005	.055	—	.120
6	—	—	.075	.010	—	.010	.075	—	—	.170
5	—	—	.040	.060	—	.060	.040	—	—	.200
4	—	—	—	.070	.030	.070	—	—	—	.170
3	—	—	—	.030	.060	.030	—	—	—	.120
2	—	—	—	—	.070	—	—	—	—	.070
1	—	—	—	—	.040	—	—	—	—	.040
Sum	.040	.070	.120	.170	.200	.170	.120	.070	.040	1.000

suppose. Psychometricians have been comfortably riding on this assumption for years, though the assumption has seldom been made explicit.

Even if transformation of variables is taken to be an appropriate way of adjusting for errors of scaling, it is not always feasible or effective. Certain kinds of censored or highly skewed distributions cannot be transformed to approximately normal form, for even the transformed distributions will continue to contain class intervals with large frequencies at the extreme, and pairs of these distributions will continue to be disparate in shape, thus lessening the possibility of rank one in the correlation matrix if Pearsonian correlations are used. This limitation also applies, I believe, to Guttman's procedures [14] for finding transformations to produce the simplest linear system for a set of variables, and Guttman's procedures will also not adequately correct for topastic error.

One solution, of course, which should be considered for the problem of scaling error is a retreat to the use of various nonparametric measures of correlation such as Spearman's rank correlation, Kendall's *tau*, and several measures which Kruskal [17] discusses in reference to 2 × 2 tables of frequencies. However, neither Spearman's rank correlation nor Kendall's *tau* will very effectively adjust for errors arising from broad grouping or censoring. The quadrant measure discussed by Kruskal will distort rank in much the same way as the phi coefficient.

If we cannot resort to nonparametric measures, I wish to call attention to the fact that there can be *parametric* measures of correlation which make no assumptions about scale, even though their interpretation involves assump-

tions about underlying distributions and regressions. Any parametric measure of correlation based on a 2×2 table is of this nature, because dichotomizing a distribution discards (censors) all information regarding scale other than ordinal. Thus, while the Pearsonian correlation coefficient may indeed in general require variables to be scaled in equal intervals, as stressed for example by Siegel ([26], p. 195), the tetrachoric r makes no such requirement (nor, for that matter, does the ϕ coefficient, but it is ruled out on other grounds). Senders was napping when she wrote in her textbook that the tetrachoric r "cannot be used for . . . ordinally scaled measurements" ([25], p. 271). Obviously, the raw, manifest data could perfectly well be ordinally scaled; the assumption of interval or ratio scaling comes in only when one is interpreting a tetrachoric r, and applies only to the underlying distribution of measurements. The tetrachoric correlation can be used in the absence of any information, or even any assumptions, about the scaling of manifest data, and can be used to adjust for the effects of scaling error.

The difficulty with the tetrachoric correlation, of course, is that it does involve reference to underlying normal bivariate surfaces with linear regressions; that is, the interpretation of r_t is meaningful to the extent that the underlying measurements conform to the model of a normal bivariate correlation surface. Let us be clear that the computation of r_t (or of r_{bis}) involves no assumptions about the data; assumptions are involved only in interpretation.

The selection of a parametric correlation measure based on a 2×2 table to correct for scaling errors therefore depends solely upon the kind of statistical model one prefers to use in the interpretation of the resulting measures. An infinite variety of models are possible, of course. It is convenient to use models which are symmetric with respect to the two marginal distributions and for which relatively simple mathematical expressions can be written. All such models will have the advantage that they assume distributions can be transformed to a common shape and therefore have the added advantage that they will not disturb rank-one conditions when they exist in submatrices of latent relationships. Actually, only one model is in common use, that is, the normal bivariate correlation surface implied by the tetrachoric correlation. Even if psychological characteristics are not distributed exactly in conformity to the normal distribution, the normal distribution is in all probability a good approximation to the true distribution—that is, to any distribution in which deviation from a central tendency becomes successively rarer as a function of the magnitude of the deviation. This alone, it seems to me, is a sufficient justification for the use of tetrachoric r's to correct for scaling errors.

But there could be still other parametric measures of correlation based upon a 2×2 table; each would involve reference to a different statistical model of the underlying correlation surface. There has been very little

investigation of such measures, and I can mention only one. It should be noted that there are two requirements for any such measure: (a) the measure (or, at least, its expected value) must be equal to the Pearsonian correlation computed from the underlying correlation surface, and (b) the magnitude of the measure must be independent of the two dichotomization points chosen to yield frequencies in the 2 $\times$ 2 table. (These requirements are met by r_t.) Obviously these requirements are not met by the phi coefficient, but it can be shown that they *are* met by the coefficient which has been symbolized as ϕ/ϕ_{max} (Cureton, [5]). It can also be shown, very easily, that this coefficient is identical to Loevinger's coefficient of homogeneity, H_t, for the case of two items (Loevinger, [19]). A truly astonishing thing about ϕ/ϕ_{max} is the nature of the underlying correlation which it implies, and for which the above requirements are met. It turns out that this underlying correlation surface is a type of bivariate rectangular surface, illustrated for the discrete case in Figure 6, such that (for $r \neq 0$) there are only two levels of frequency: the frequencies in diagonal cells manifest one uniform level of frequency, and the frequencies in nondiagonal cells manifest another level of frequency. Figure 6, for example, shows the underlying correlation surface which is characterized by a Pearsonian correlation coefficient equal to .50 and which will yield ϕ/ϕ_{max} = .50 for all possible discrete dichotomization points indicated. Similar surfaces can be constructed for other values of ϕ/ϕ_{max}.

This kind of correlation surface seems just a bit improbable, and I believe most of us would shy away from it as a model for interpreting correlation measures. If so, the use of ϕ/ϕ_{max} and consequently of Loevinger's

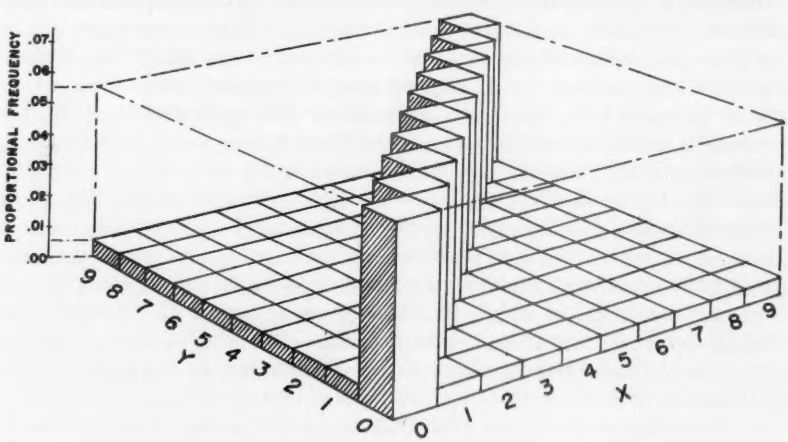

FIGURE 6

Illustrating the Type of Bivariate Rectangular Distribution for which ϕ/ϕ_{max} is Constant and Equal to the Pearsonian r for the Distribution as a Whole (r = .5)

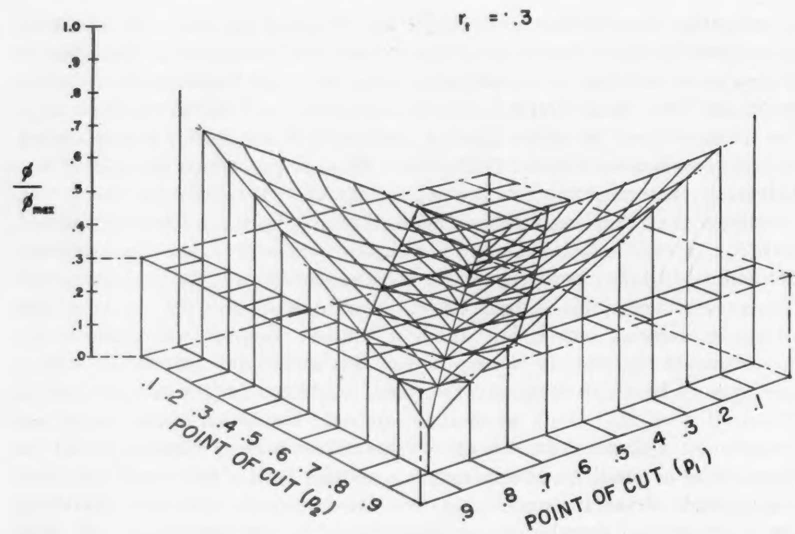

FIGURE 7

Values of ϕ/ϕ_{max} Obtained at Various Pairs of Cutting Points from a Normal Bivariate Correlation Surface, $r = .3$

H_t becomes much less desirable than might have been thought. As we bid farewell to ϕ/ϕ_{max} and H_t, however, it may be instructive to inquire how much difference it makes whether we use r_t or ϕ/ϕ_{max}. Suppose we apply ϕ/ϕ_{max} to a normal bivariate surface; with various dichotomization points, how well will the resulting values approximate the uniform values of r_t which would be yielded by those dichotomizations? The answer is that the discrepancies are rather considerable. Figures 7 and 8 show the surfaces of ϕ/ϕ_{max} coefficients which result for correlation parameters of .30 and .80. Note that when the dichotomization points are equal or nearly equal, the ϕ/ϕ_{max} coefficients are *less* than the corresponding tetrachoric r's and when they become decidedly unequal, the ϕ/ϕ_{max} coefficients are increasingly *greater* than the tetrachoric r's, as represented by the flat surfaces cutting through the curved surfaces. Let us say goodby, and not *au revoir*, to ϕ/ϕ_{max}. Incidentally, our findings with regard to ϕ/ϕ_{max} should serve as a warning against computing r/r_{max} for the more general case, e.g., for distributions as disparate as those of Figure 1.

Even though there is no final justification for the use of the tetrachoric correlation to adjust for scaling errors in psychometric data, the successful avoidance of "difficulty factors" and the general orderly appearance of the factorial results even in highly disorderly sets of raw data lends support to

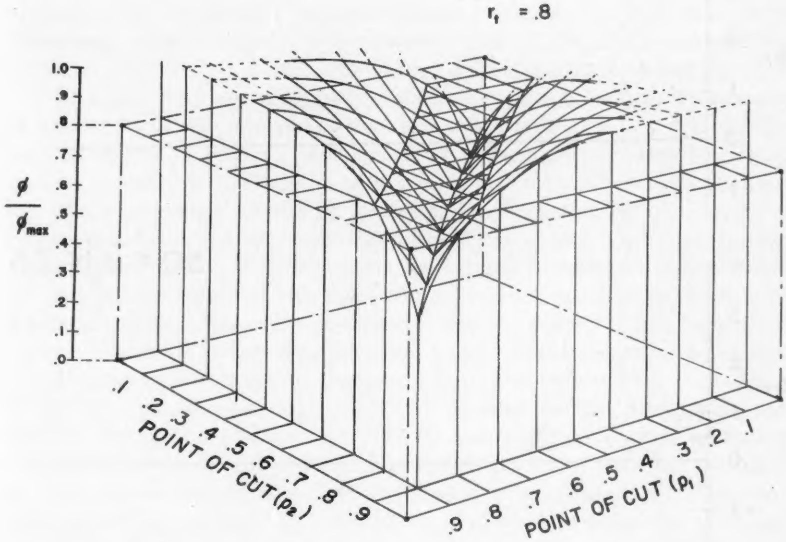

FIGURE 8
As in Figure 7, but for $r = .8$

this technique. Partly to demonstrate this, I have prepared Figure 9 on the basis of data from the aphasia study mentioned earlier. A random selection of 20 variables was made; skewness coefficients ($g_1 = \mu_3/\sigma^3$) were computed for each distribution. Matrices of Pearsonian and of tetrachoric correlations were then computed. The *algebraic* difference Δ_r between each tetrachoric coefficient and its Pearsonian counterpart was then plotted against the *absolute* difference Δ_g between the skewness coefficients of the respective distributions. The plots are made for four levels of the tetrachoric coefficients, on the assumption that the tetrachoric correlations more closely approximate measures of the true latent relationships among the variables. (The plots would be less orderly if they had been arranged in accordance with the level of Pearsonian r—and this fact is perhaps another justification for the use of the tetrachoric r.) It can be seen that as the value of the tetrachoric r increases, the boost from the Pearsonian value becomes more and more dependent upon the discrepancies of the marginal distributions as measured—inadequately, to be sure—by the differences in their skewnesses. The differences are, in fact, not at all large until the size of the latent relationship—as estimated by the tetrachoric r—becomes appreciable.

Furthermore, after factor analysis of the results in the complete 69 × 69 matrix it was possible to assemble large clusters of variables which exhibited

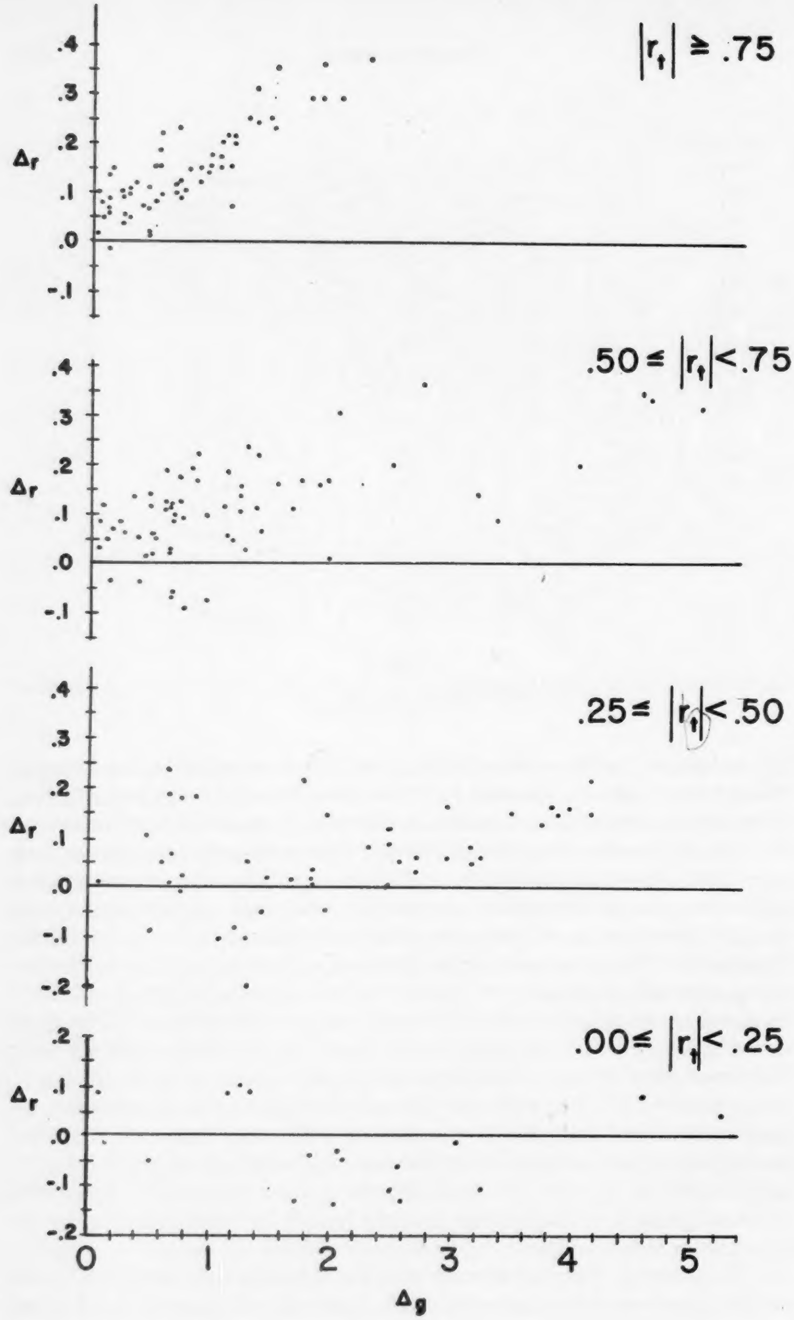

FIGURE 9

Plots of Δ_g vs. Δ_r, from Empirical Data, for Various Levels of r_t (See explantion in text)

approximately hierarchical intercorrelations, despite the fact that their skewnesses were extremely heterogeneous. Use of Pearsonian correlations for these same clusters would have yielded a rank greater than one.

Although there was undoubtedly scedastic error in these data, the amount of topastic error was minimal because few of the tests gave any opportunity for the subject to obtain a score point by mere guessing. Therefore, there was no necessity to attempt to adjust the joint frequency distributions for the effects of chance success by guessing. In effect, the use of tetrachoric correlations in the aphasia study was primarily an attempt to control for the scaling error, which was evidenced by the marked disparity of distributions.

Let me now speak of two cases where danger of spurious results due to failure to adjust for the effects of topastic error was present. First, Dingman's study [7] deserves further consideration. I have already mentioned the fact that Dingman failed to use the proper procedure for correction for the effects of topastic error, but he could claim, at least after the fact, that he did not need to because he obtained no difficulty factor strong enough to distort content factors. It is probable that Dingman's battery of tests was too small to allow him to obtain clear difficulty factors; also, clustering of the tests with respect to content was apparently sufficiently marked to minimize any tendencies for the tests to cluster in terms of difficulty level. Dingman does not present data which would allow one to judge how much disparity existed between the test score distributions, but I suspect that the disparity was in no case marked. Even so, Dingman did obtain a factor which he was willing to call a difficulty factor; the loadings show a general trend in the direction of an association with the difficulty level of the tests, particularly when Pearsonian r's were used. The small size of the loadings is not out of line with the amount of perturbation that one would expect in view of the nature of these data. In short, the results are not out of line with the formulations of Ferguson and of Carroll.

Second, I wish to make further comments about the study by Guilford [13], which prompted Gourlay's article [12]. It was Guilford's article that also prompted *my* interest in the question of spurious difficulty factors resulting from the improper choice of correlation measures. It seemed to me highly unlikely, in view of all our known results in psychophysics, that there could be separate factors in pitch discrimination ability for different levels of difficulty, that is, differences in pitch. If a person could perform well in discriminating very small differences in pitch, he certainly could do well with large differences in pitch, unless he were utterly bored, or misunderstood the task; conversely, a person who failed to discriminate large differences would certainly fail to discriminate small differences in pitch. I therefore began scrutinizing Guilford's data with the idea that his factors were artifacts of the statistical methods employed, and discovered that contrary to opinions which were then current, even the tetrachoric correla-

tion coefficient was subject to biasing effects of difficulty level when there was a possibility of chance success by guessing. In the original manuscript of a paper submitted to *Psychometrika* in 1944 I suggested how Guilford's results could be explained as due to statistical artifacts, but the editors did not consider my demonstration sufficiently well worked out, and I was forced to agree; hence, the published paper [3] omitted any consideration of Guilford's data. Part of the difficulty was that I did not have available the raw joint distributions of the subtests of the Seashore Sense of Pitch test which Guilford had analyzed and could not show exactly how his results could have been predicted from my theoretical developments; consequently, I proceeded to administer this test to a large number of students in order to accumulate the necessary data. In 1950 I presented some tentative results at the meetings of the Psychometric Society, but because I was still not satisfied with the results I did not publish them. It was about this time that Gourlay [12] published his article pointing out the artifact in Guilford's results; it may be noted, incidentally, that although Gourlay made reference to my work his demonstration was at a more elementary level than would have been possible if he had made full use of my formulations, because his demonstration relied solely on expected tetrachoric correlations between *single* items rather than between sets of items of equal difficulty.

In recent years, I have been able to develop a model to account more completely for Guilford's results, that is, a model allowing for both scedastic and topastic variation. There is no space here for explicating this; I may say, however, that some of the figures presented in this paper are based on my work with data comparable to Guilford's. Figure 2, for example, is the underlying perfect relationship that I assume between subtests B and I of the Seashore Sense of Pitch test if there were neither scedastic nor topastic error in the data, and if it were possible to measure the individual's psychophysical limen perfectly by means of either test. However, because in actuality each test contains only ten items, and the tests are at different levels of difficulty, the true scores would at best be subject to censoring, particularly at the extremes of the distribution. This censoring is depicted in the labeling of the coordinates of Figure 2; Figure 1 and Table 1 represent the joint distribution of the true raw scores thus censored. Table 2 shows the expected joint distribution if there were only scedastic error, and Table 3, as we have said, shows the expected joint distribution if topastic error (with $c = .5$) is also assumed to be operating uniformly for all items. The expected Pearsonian $r = .3160$ and the expected tetrachoric $r_t = .56$. The corresponding observed values *in my data* ($N = 1082$) are as follows: $r = .228$; $r_t = .40$. Evidently the Seashore test data do not quite conform to the perfect relationship assumed in Figure 2, but since among all the subtests the obtained correlations are highly related, linearly, to those expected, we may take it that these subtests indeed measure one and only one trait, and that it is ade-

quately demonstrated that Guilford's "difficulty factors" are artifacts of the kinds of correlations used. Guilford could have made a proper factor analysis of the Seashore Sense of Pitch test only by using tetrachoric correlations based on joint distributions corrected at least for the effects of topastic error.

The case of Guilford's difficulty factors affords a prime example of a situation where the nature of the data must be carefully considered before choosing a statistical method. And it is not merely the superficial appearance of the data that must be considered, but also the conditions under which they were obtained and the possible models for accounting for them. These matters, I think, are properly within the purview of the Psychometric Society.

Appendix A

A Method for Finding the Limits of the Product-Moment Correlation Coefficient for Any Two Distributions

1. Set up the two frequency distributions with the same orientation, e.g., with the highest or "best" scores at the top.

2. In each distribution, obtain the cumulative proportional frequencies for each score or class interval, starting the cumulation at the bottom. The last cumulative proportion is unity; drop it from further computations.

3. The cumulative proportions will now be denoted k_x and k_y for the two distributions respectively. Find $\sum k_x$ and $\sum k_y$.

4. Assign integers ($n_a = 0, 1, 2, 3, \cdots$) to each cumulative proportion starting from the top (excluding the one dropped in step 2). Obtain $\sum n_a k_x$ and $\sum n_a k_y$.

Finding Maximum Correlation — Example

Distribution of X					Distribution of Y					Rearrangement of Values of k	
X	f	p	k_x	n_a	Y	f	p	k_y	n_a	k_t	n_a
4	8	.16	(1.00)	–	4	4	.08	(1.00)	–	.92	0
3	12	.24	.84	0	3	10	.20	.92	0	.84	1
2	15	.30	.60	1	2	20	.40	.72	1	.72	2
1	10	.20	.30	2	1	8	.16	.32	2	.60	3
0	5	.10	.10	3	0	8	.16	.16	3	.32	4
										.30	5
N =	50	1.00	$\Sigma k_x = 1.84$			50	1.00	$\Sigma k_y = 2.12$		.16	6
										.10	7
			$\Sigma n_a k_x = 1.50$					$\Sigma n_a k_y = 1.84$			
										$\Sigma n_a k_t = 8.52$	

$$r_{\max} = \frac{8.52 - 1.50 - 1.84 - (1.84)(2.12)}{[1.84 + 2(1.50) - (1.84)^2]^{1/2}\,[2.12 + 2(1.84) - (2.12)^2]^{1/2}} = .9283$$

5. Arrange the cumulative proportions k from the two distributions in a new single list, in order of decreasing magnitude. Call entries in this list k_t and assign integers ($n_a = 0, 1, 2, 3, \cdots$) to the entries in order starting from the top. (Assign separate numbers even if two or more values are identical.) Obtain $\sum n_a k_t$.

6. The maximum positive product-moment correlation is found by evaluating the formula

$$r = \frac{\sum n_a k_t - \sum n_a k_x - \sum n_a k_y - (\sum k_x)(\sum k_y)}{\sqrt{[\sum k_x + 2 \sum n_a k_x - (\sum k_x)^2][\sum k_y + 2 \sum n_a k_y - (\sum k_y)^2]}}.$$

7. To find the maximum negative r, reverse the orientation of one of the distributions and repeat the procedure.

Appendix B

Correction of Joint Distribution for Topastic Error and Determination of a Non-Topastically Affected Tetrachoric r

1. It is first necessary to estimate distributions of non-topastically affected scores. One method is the following, although it is not very satisfactory because the results often contain negative frequencies. Let $\mathbf{L}$ be a row vector containing the $(n + 1)$ frequencies of non-topastically affected scores $L = 0, 1, 2, \cdots, n$, where n is the number of items, and let $\mathbf{C}$ be the corresponding row vector containing the frequencies of topastically affected scores $C = 0, 1, 2, \cdots, n$ (that is, the obtained scores). Assume that the topastic probability is uniform for all items and equal to c; d is the complement of c. Then the relation between $\mathbf{L}$ and $\mathbf{C}$ is expressed by the formula

$$\mathbf{L}\mathbf{T}_{n,c} = \mathbf{C}, \tag{1}$$

where $\mathbf{T}_{n,c}$ is a square matrix of order $(n + 1)$ with the general form

		C					
		0	1	2	$\cdots$	$(n-1)$	n
	0	d^n	$nd^{n-1}c$	$\frac{1}{2}n(n-1)\,d^{n-2}c^2$	$\cdots$	ndc^{n-1}	c^n
	1	—	d^{n-1}	$nd^{n-2}c$	$\cdots$	ndc^{n-2}	c^{n-1}
	2	—	—	d^{n-2}	$\cdots$	ndc^{n-3}	c^{n-2}
L	3	—	—	—	$\cdots$	ndc^{n-4}	c^{n-3}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	$n-1$	—	—	—	$\cdots$	d	c
	n	—	—	—	$\cdots$	—	1

That is, each row of $\mathbf{T}_{n,c}$ contains, for its last $(n - L + 1)$ entries, the expansion of the binomial $(d + c)^{n-L}$.

2. For any distribution, the vector $\mathbf{L}$ can then be estimated by solving (1):

$$\mathbf{L} = \mathbf{C}\mathbf{T}_{n,c}^{-1} . \tag{2}$$

3. Estimate by formula (2) or by other methods the vector of non-topastically affected scores for each of the variables, and find the dichotomization point nearest the median of this vector. (It may be useful to choose several pairs of dichotomization points and carry out the subsequent steps for each, in order to assess the variance of estimate.)

4. For each distribution, let q be the obtained proportion below a dichotomy and $\bar{k}$ the estimated non-topastic proportion below this point. Then find $\bar{c}$, the topastic probability for this dichotomy, by solving

$$\bar{k} = q/(1 - \bar{c}). \tag{3}$$

5. Construct a 2×2 table for the two distributions corrected for the effects of topastic error. The marginal proportions will have been computed in step (4). The corrected proportion below *both* dichotomization points is found as

$$\bar{k}_{xy} = q_{xy}/(1 - \bar{c}_x)(1 - \bar{c}_y), \tag{4}$$

where q_{xy} is the obtained proportion falling below both dichotomies.

6. Compute the topastically corrected tetrachoric r from the completed 2×2 table.

REFERENCES

[1] Binder, A. Consideration of the place of assumptions in correlational analysis. *Amer. Psychologist*, 1959, **14**, 504-510.

[2] Carroll, J. B. A factor analysis of verbal abilities. *Psychometrika*, 1941, **6**, 279-307.

[3] Carroll, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, **10**, 1-19.

[4] Cattell, R. B. *Factor analysis*. New York: Harper, 1952.

[5] Cureton, E. E. Note on ϕ/ϕ_{max}. *Psychometrika*, 1959, **24**, 89-91.

[6] DeSoto, C. B. and Kuethe, J. L. On the relation between two variables. *Educ. psychol. Measmt*, 1960, **20**, 743-749.

[7] Dingman, H. F. The relation between coefficients of correlation and difficulty factors. *Brit. J. statist. Psychol.*, 1958, **11**, 13-17.

[8] Dunlap, J. W. Psychometrics—a special case of the Brahman theory. *Psychometrika*, 1961, **26**, 65-71.

[9] Ferguson, G. A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, **6**, 323-329.

[10] Furfey, P. H. Comment on "The needless assumption of normality in Pearson's r." *Amer. Psychologist*, 1958, **13**, 545-546.

[11] Gage, N. L. and Damrin, D. E. Reliability, homogeneity, and number of choices. *J. educ. Psychol.*, 1950, **41**, 385-404.

[12] Gourlay, N. Difficulty factors arising from the use of tetrachoric correlations in factor analysis. *Brit. J. statist. Psychol.*, 1951, **4**, 65-73.
[13] Guilford, J. P. The difficulty of a test and its factor composition. *Psychometrika*, 1941, **6**, 67-77.
[14] Guttman, L. Metricizing rank-ordered or unordered data for a linear factor analysis. *Sankhyā*, 1959, **21**, 257-268.
[15] Heath, H. A. An empirical study of correlation involving a half-normal distribution. *Psychol. Rep.*, 1961, **9**, 85-86.
[16] John, E. and Burt, C. A reply to Mr. Gourlay's criticisms. *Brit. J. statist. Psychol.*, 1951, **4**, 73-76.
[17] Kruskal, W. H. Ordinal measures of association. *J. Amer. statist. Ass.*, 1958, **53**, 814-861.
[18] LaForge, R. Comment on "The needless assumption of normality in Pearson's *r*." *Amer. Psychologist*, 1958, **13**, 546.
[19] Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, **61**(4), 1-49.
[20] Luce, R. D. On the possible psychophysical laws. *Psychol. Rev.*, 1959, **66**, 81-95.
[21] Milholland, J. E. Comment on "The needless assumption of normality in Pearson's *r*." *Amer. Psychologist*, 1958, **13**, 544-545.
[22] Nefzger, M. D. and Drasgow, J. The needless assumption of normality in Pearson's *r*. *Amer. Psychologist*, 1957, **12**, 623-625.
[23] Norris, R. C. and Hjelm, H. F. Non-normality and product-moment correlation. *J. exp. Educ.*, 1961, **29**, 261-270.
[24] Schuell, H. and Jenkins, J. J. The nature of language deficit in aphasia. *Psychol. Rev.*, 1959, **66**, 45-67.
[25] Senders, V. L. *Measurement and statistics*. New York: Oxford, 1958.
[26] Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
[27] Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1-49.
[28] Thomson, G. H. *The factorial analysis of human ability*. (5th ed.) Boston: Houghton-Mifflin, 1951.
[29] Thurstone, L. L. Primary mental abilities. *Psychometric Monogr. No. 1*. Chicago: Univ. Chicago Press, 1938.
[30] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.

Manuscript received 9/6/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

STOCHASTIC LEARNING THEORIES FOR A RESPONSE CONTINUUM WITH NON-DETERMINATE REINFORCEMENT*

PATRICK SUPPES, STANFORD UNIVERSITY

AND

JOSEPH L. ZINNES, INDIANA UNIVERSITY

Continuous analogues of the finite linear and stimulus sampling theories are developed for non-determinate reinforcement schedules. Closed-form expressions are derived for the asymptotic response distribution and certain sequential statistics. Computations for a target experiment are given to illustrate the character of the theoretical results.

Recent extensions of the stochastic learning theories [1, 2] to a continuum of responses have dealt exclusively with determinate reinforcement schedules. The latter refer to experimental tasks in which the subject is informed of the correct response on each trial or, in the case of animal subjects, to experiments using a correction procedure. In the present paper further extension to the non-determinate case is considered.

Although the theory to be developed here is intended to have some generality, it will be helpful to have in mind one of the experiments underway at present (hereafter called the *Target Experiment*). In this experiment subjects are instructed to locate or "hit" an unseen target which is said to be located at some point on the circumference of a circle. (For a description of the experimental apparatus, see [3].) The exact position of the target on each trial is determined by sampling from a fixed distribution defined over the circumference. If the subject's response lies within a specified distance of the target he is informed that he has a hit on that trial, otherwise a miss. Since the subject is free to choose, at least theoretically, any point on the circumference of the circle, the response alternatives may be said to lie on a continuum. The non-determinate aspect of the experiment refers to the fact that the subject is not informed of the exact location of the target after a miss (or a hit).

The plan of this paper is to develop separately the general equations for two types of theories—continuous analogues of the finite linear and stimulus sampling theories—and to illustrate more specifically the character of the models by applying them to the Target Experiment. In these examples

*This research was supported by the Rockefeller Foundation and the Group Psychology Branch of the Office of Naval Research.

we shall use some arbitrary parameter values, although in any actual application of the model, these parameters would have to be estimated from the data.

The Linear Theory

The extension of the continuous linear theory to non-determinate conditions is more easily described by considering briefly the determinate case. Denoting the value of the response random variable on trial n by x_n and the value of the reinforcement variable by y_n ($a \leq x_n \leq b$ and $a \leq y_n \leq b$) the sequence of experimental outcomes preceding the $(n + 1)$st trial is described by the $2n$-dimensional vector

$$s_n = (y_n, x_n, y_{n-1}, x_{n-1}, \cdots, y_1, x_1). \tag{1}$$

The response distribution on trial $n + 1$, which is of experimental interest, can then be defined as the marginal distribution obtained by integrating over the $2n$ dimensions. In particular, if $j_{n+1}(x, y_n, x_n, \cdots, y_1, x_1)$ denotes the joint density function of the first $n + 1$ responses and n reinforcements, then the response density of the $(n + 1)$st response $r_{n+1}(x)$ is defined as

$$r_{n+1}(x) = \int_a^b \cdots \int_a^b j_{n+1}(x, y_n, x_n, \cdots, y_1, x_1)\, dy_n\, dx_n \cdots dy_1\, dx_1, \tag{2}$$

which for simplicity is written as

$$r_{n+1}(x) = \int_a^b j_{n+1}(x, s_n)\, ds_n. \tag{3}$$

In (2) it has been assumed that both reinforcement and response variable[8] are continuous variables. Although under determinate reinforcement conditions one could assume a discrete reinforcement variable (and hence a discrete reinforcement distribution), it is more natural or perhaps merely more interesting to assume that both the response and reinforcement variables are continuous. These conditions permit the usual one-to-one correspondence between response alternatives and the set of reinforcing events. For non-determinate conditions, on the other hand, simplicity is obtained by confining the reinforcement variable to two values: 1 denoting a correct or rewarded response and 0 an incorrect or unrewarded response. (More complicated non-determinate reinforcement schedules can be obtained by giving the subject additional information on incorrect trials.)

Using Y_n (or on occasion $Y_{i,n}$, where $Y_{0,n} = 0$ and $Y_{1,n} = 1$) as the discrete reinforcement variable the sequence of experimental outcomes s_n becomes

$$s_n = (Y_n, x_n, Y_{n-1}, x_{n-1}, \cdots, Y_1, x_1),$$

and the response density $r_n(x)$, as defined in (2), then involves summing over n dimensions and integrating over the remaining n dimensions. (We

shall continue, however, to employ the terminology of (3) for this case as well.)

As in the finite linear theories, the basic assumptions of the continuous theories are stated recursively, that is, as rules or laws indicating how the response densities (instead of response probabilities) change after each trial. There are two possible outcomes on each trial and hence two recursions to consider. If the response on trial n is correct (i.e., $Y_n = 1$), then it is assumed that

$$j_{n+1}(x \mid Y_{1,n}, x_n, s_{n-1}) = (1 - \theta)j_n(x \mid s_{n-1}) + \theta k_h(x, x_n). \tag{4}$$

The last term in (4), $k_h(x, x_n)$, requires some comment. In general, it will be assumed that this function is unimodal and symmetric about x_n so that the effect of this term is to spread out or generalize the reinforcement effects to neighboring points on the response continuum. To assume that the reinforcement is concentrated on the continuum at the point of the reinforced response is psychologically untenable—the subject cannot discriminate this well—and furthermore it leads to mathematically untractable expressions. It is perhaps simplest to think of the right side of (4) as merely involving the weighting of two distributions, one distribution of some complicated character reflecting the subject's past history, and the other a distribution which is unimodal and symmetric about the response reinforced on the given trial. The symmetric aspect of the k_h distribution could of course be questioned when end effects are present, but it is reasonably compelling for circular or periodic continua.

When the response on trial n is incorrect we have an analogous expression:

$$j_{n+1}(x \mid Y_{0,n}, x_n, s_{n-1}) = (1 - \theta)j_n(x \mid s_{n-1}) + \theta k_m(x, x_n). \tag{5}$$

The exact nature of the k_m function is not as obvious. Two possibilities will be considered in the application section following.

$$\text{(i)} \qquad k_m = \frac{1}{b - a}$$

and

$$\text{(ii)} \qquad k_m = k_h(x, x_n - \pi).$$

Assumption (i), the assumption of a uniform distribution, has the effect of "flattening out" the distribution based on the previous trials when a miss occurs, while assumption (ii) treats reinforcement and nonreinforcement effects in a complementary manner. The maximum point of the k_h distribution corresponds to the minimum point of the k_m distribution and similarly for the minimum point of the k_h distribution. More detailed properties are discussed in the following section.

By combining (4) and (5) with (3) a simple recursion involving the

response density $r_n(x)$ may be obtained. The details of the derivation follow the procedure given in [1], so that here we need merely sketch the argument. Equation (3) can be written as follows, where any obvious limits of integration are omitted:

$$r_{n+1}(x) = \iint \sum_i j_{n+1}(x, Y_{i,n}, x_n, s_{n-1})\, dx_n\, ds_{n-1}, \tag{6}$$

or more explicitly as

$$\begin{aligned} r_{n+1}(x) = &\iint j_{n+1}(x, Y_{0,n}, x_n, s_{n-1})\, dx_n\, ds_{n-1} \\ &+ \iint j_{n+1}(x, Y_{1,n}, x_n, s_{n-1})\, dx_n\, ds_{n-1}. \end{aligned} \tag{7}$$

Consider the second term on the right side of (7); rewriting this term using conditional probabilities

$$\iint j_{n+1}(x \mid Y_{1,n}, x_n, s_{n-1}) j(Y_{1,n} \mid x_n, s_{n-1}) j(x_n \mid s_{n-1}) j(s_{n-1})\, dx_n\, ds_{n-1} \tag{8}$$

and substituting (4) into (8) gives

$$\begin{aligned} &\iint [(1 - \theta) j_n(x \mid s_{n-1}) \\ &\qquad + \theta k_h(x, x_n)] j(Y_{1,n} \mid x_n, s_{n-1}) j(x_n \mid s_{n-1}) j(s_{n-1})\, dx_n\, ds_{n-1} \\ &= \iint (1 - \theta) j_n(x \mid s_{n-1}) j(s_{n-1}) j(Y_{1,n} \mid x_n, s_{n-1}) j(x_n \mid s_{n-1})\, dx_n\, ds_{n-1} \\ &\qquad + \iint \theta k_h(x, x_n) j(Y_{1,n} \mid x_n, s_{n-1}) j(x_n, s_{n-1})\, dx_n\, ds_{n-1}. \end{aligned} \tag{9}$$

In a similar manner (5) may be substituted into the first term on the right side of (7) to give

$$\begin{aligned} &\iint (1 - \theta) j_n(x \mid s_{n-1}) j(s_{n-1}) j(Y_{0,n} \mid x_n, s_{n-1}) j(x_n \mid s_{n-1})\, dx_n\, ds_{n-1} \\ &\qquad + \iint \theta k_m(x, x_n) j(Y_{0,n} \mid x_n, s_{n-1}) j(x_n, s_{n-1})\, dx_n\, ds_{n-1}. \end{aligned} \tag{10}$$

The expressions in (9) and (10) may be combined and simplified by noting first of all that since

$$j(Y_{0,n} \mid x_n, s_{n-1}) + j(Y_{1,n} \mid x_n, s_{n-1}) = 1$$

the first terms of (9) and (10) can be combined and the integration over x_n carried out. Secondly, since it is assumed, in accordance with the usual

non-determinate reinforcement schedules, that Y_n does not depend on s_{n-1}, that is, for example,

$$j(Y_{0,n} \mid x_n, s_{n-1}) = j(Y_{0,n} \mid x_n),$$

the integration over s_{n-1} in each of the second terms in (9) and (10) can be performed. Thus combining these terms

$$\begin{aligned}(11) \qquad r_{n+1}(x) = (1 - \theta) \int j_n(x \mid s_{n-1}) j(s_{n-1})\, ds_{n-1} \\ + \theta \int k_h(x, x_n) j(Y_{1,n} \mid x_n) j(x_n)\, dx_n \\ + \theta \int k_m(x, x_n) j(Y_{0,n} \mid x_n) j(x_n)\, dx_n .\end{aligned}$$

Equation (11) can be simplified further by replacing the first integral with $r_n(x)$ and to conform with the usual notation in the finite theories we shall replace $j(Y_{1,n} \mid x_n)$ with $\pi_{1,n}(x_n)$ and $j(Y_{0,n} \mid x_n)$ with $\pi_{0,n}(x_n)$; $j(x_n)$ is, of course, nothing but $r_n(x)$. Thus we have for the response density the simple recursion

$$\begin{aligned}(12) \qquad r_{n+1}(x) = (1 - \theta) r_n(x) + \theta \int k_h(x, x_n) \pi_{1,n}(x_n) r_n(x_n)\, dx_n \\ + \theta \int k_m(x, x_n) \pi_{0,n}(x_n) r_n(x_n)\, dx_n .\end{aligned}$$

From (12) the equation for the asymptotic response density $r(x)$ follows directly:

$$(13) \qquad r(x) = \int k_h(x, x_n) \pi_1(x_n) r(x_n)\, dx_n + \int k_m(x, x_n) \pi_0(x_n) r(x_n)\, dx_n .$$

(The existence of this asymptotic response density and of other asymptotic quantities considered below can be established by relatively direct but lengthy arguments, so that we shall not enter into them in this paper.) Equation (13) is a linear homogeneous integral equation in $r(x)$. The ease of solving (13) explicitly depends on the form of the functions k_h, k_m, and π_1. Note that the appearance of x_n in (13) does not mean that $r(x)$ depends on n, for x_n is the variable of integration.

The functions π_1 and π_0 in (13) are determined by the experimenter so they can be considered as known functions here. In the Target Experiment these functions are determined indirectly. If the exact position of the target is denoted by y and its density by $f(y)$ then the probability of a hit on trial n given response x_n, i.e., $\pi_1(x_n)$, is given in terms of $f(y)$ by

$$\pi_1(x_n) = \int_{x_n - a}^{x_n + a} f(y)\, dy,$$

where α, another experimenter-determined parameter, specifies the permissible range about the target or 2α the effective size of the target.

Special Case: Asymptotic Distributions in the Target Experiment

In this section we shall take certain one-parameter functions for k_h, k_m, and π_1 [or $f(y)$] and investigate their implications for the Target Experiment.

For the functions $k_h(x, x_n)$ and $f(y)$ we assume

$$k_h(x, x_n) = C_j \cos^{2j}\left(\frac{x - x_n}{2}\right) \qquad (j = 1, 2, \cdots), \tag{15}$$

$$f(y) = C_i \cos^{2i}\left(\frac{y}{2}\right) \qquad (i = 1, 2, \cdots), \tag{16}$$

where C_i (and C_j), the normalization factor, is given by

$$C_i = \frac{2^{(2i-1)}(i!)^2}{\pi(2i)!}. \tag{17}$$

Some comment on the use of the cosine function raised to even powers as a density function is perhaps called for here. This function has three specific properties which make it particularly convenient to use in this context: the function is periodic, the indefinite integral has a closed-form expression, and, most importantly, it leads to a degenerate kernel in (13). [The kernel refers to the terms in each integral of (13) excluding the function $r(x)$. It is said to be degenerate if the two variables, x and x_n, can be separated; that is, for example, if functions g_i and h_i exist such that for some N

$$\pi_1(x_n)k_n(x, x_n) = \sum_{i=1}^{N} g_i(x)h_i(x_n). \tag{18}$$

The cosine function in (16) has this property since by the usual trigonometric identities

$$\cos^{2i}\left(\frac{x - x_n}{2}\right) = \sum_{j=0}^{i} a_j \cos jx \cos jx_n + \sum_{j=1}^{i} b_j \sin jx \sin jx_n, \tag{19}$$

where a_j and b_j are constants independent of x and x_n.] These three properties produce considerable simplification in solving (13) for $r(x)$.

Despite their somewhat unconventional appearance both density functions in (15) and (16) lead to reasonably conventional—we might almost say normal—distributions in the interval $(-\pi, \pi)$ and $(x - \pi, x + \pi)$, respectively. The distributions are unimodal in this interval, symmetric, and have two inflection points for $j > 1$ placed symmetrically about the mean. The value of the exponent of the cosine in each case determines the variance of the distribution. For example, when $j = 2$, the variance equals .79, and,

to compare this particular distribution with the normal distribution, .67 of the area lies within one sigma of the mean.

For the function $k_m(x, x_n)$ there are two possibilities (described previously) which we shall consider. They lead to what we term the *Uniform Theory* (or U theory) and the *Symmetric Theory* (or S theory).

Uniform Theory

This theory is characterized by the assumption

$$k_m(x, x_n) = \frac{1}{2\pi}. \tag{20}$$

For illustrative purposes in this section it will be assumed that $j = 1$ in (15), i.e.,

$$k_h = \frac{1}{\pi} \cos^2 \left(\frac{x - x_n}{2}\right), \tag{21}$$

although previous experimental work [3] indicates that a more realistic assumption would involve a distribution with a variance approximately equal to .58 and this corresponds approximately to $j = 6$. For the reinforcement function $f(y)$ we shall take

$$f(y) = \frac{1}{\pi} \cos^2 \left(\frac{y}{2}\right), \tag{22}$$

and therefore

$$\pi_1(x_n) = \frac{1}{\pi} \int_{x_n - \alpha}^{x_n + \alpha} \cos^2 \frac{y}{2}\, dy = \frac{1}{\pi} (\alpha + \sin \alpha \cos x_n). \tag{23}$$

Now to solve (13) for $r(x)$. Since the method of solving integral equations with degenerate kernels may not be familiar to the reader the details of the solution will be given. Some initial simplification of (13) is obtained by substituting (15) into (13), replacing $\pi_0(x_n)$ with $1 - \pi_1(x_n)$, and letting

$$G = \frac{1}{2\pi} \int_{-\pi}^{\pi} \pi_1(x_n) r(x_n)\, dx_n, \tag{24}$$

where G is a constant independent of x. Thus (13) reduces to

$$r(x) = \frac{1}{2\pi} - G + \int_{-\pi}^{\pi} k_h(x, x_n) \pi_1(x_n) r(x_n)\, dx_n. \tag{25}$$

Consider the integral term in (25). First expanding (21)

$$k_h(x, x_n) = \frac{1}{\pi} \cos^2 \left(\frac{x - x_n}{2}\right) = \frac{1}{2\pi} (1 + \cos x \cos x_n + \sin x \sin x_n) \tag{26}$$

and then using (26) and (23) we have for this integral,

$$(27)\quad \int k_h(x, x_n)\pi_1(x_n)r(x_n)\,dx_n$$

$$= \int \frac{1}{2\pi}(1 + \cos x \cos x_n + \sin x \sin x_n)\,\frac{1}{\pi}(\alpha + \sin\alpha \cos x_n)r(x_n)\,dx_n$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \pi_1(x_n)r(x_n)\,dx_n + \frac{1}{2\pi^2}\int_{-\pi}^{\pi} [(\alpha \cos x)\cos x_n$$

$$+ (\sin\alpha \cos x)\cos^2 x_n + (\alpha \sin x)\sin x_n$$

$$+ (\sin\alpha \sin x)\sin x_n \cos x_n]r(x_n)\,dx_n\,,$$

which simplifies to

$$(28)\quad \int k_h(x, x_n)\pi_1(x_n)r(x_n)\,dx_n = G + A\cos x + B\sin x,$$

where the coefficients A and B, representing the definite integrals over x_n, are independent of x. Substituting (28) into (25) indicates that the density $r(x)$ must have the form

$$(29)\quad r(x) = \frac{1}{2\pi} + G - (G + A\cos x + B\sin x)$$

$$= \frac{1}{2\pi} + A\cos x + B\sin x.$$

The coefficients in (29), A and B, are evaluated by obtaining a set of linear simultaneous equations as follows. The expression for $r(x)$ in (29) is placed into (27), replacing, first of all, x by x_n, and the resulting integrals evaluated. This gives a second expression for $r(x)$, viz.,

$$(30)\quad r(x) = \frac{1}{2\pi} + \frac{1}{2\pi^2}[\tfrac{1}{2}\sin\alpha\cos x + A\pi\alpha\cos x + B\pi\alpha\sin x]$$

$$= \frac{1}{2\pi} + \left(\frac{\sin\alpha}{4\pi^2} + \frac{A\alpha}{2\pi}\right)\cos x + \left(\frac{B\alpha}{2\pi}\right)\sin x.$$

Since both (29) and (30) must hold for all values of x we may equate the coefficients of $\cos x$ and $\sin x$ and thus obtain the equations

$$(31)\quad \begin{cases} A = \left(\dfrac{\sin\alpha}{4\pi^2} + \dfrac{A\alpha}{2\pi}\right) \\ B = \dfrac{B\alpha}{2\pi} \end{cases}$$

which give

(32)
$$\begin{cases} A = \dfrac{\sin \alpha}{2\pi(2\pi - \alpha)} \\ B = 0, \end{cases}$$

so that finally

(33)
$$r(x) = \frac{1}{2\pi}\left(1 + \frac{\sin \alpha}{2\pi - \alpha} \cos x\right).$$

The asymptotic response distribution indicated by (33) is, as expected, unimodal and symmetric about $x = 0$. Thus the means of the reinforcement distribution and response distribution coincide although the two distributions differ in variance. The variance of the response distribution, obtained directly from (33), is equal to

(34)
$$\sigma_u^2 = \frac{\pi^2}{3} - 2\left(\frac{\sin \alpha}{2\pi - \alpha}\right),$$

and the variance of the reinforcement distribution $f(y)$ equals

(35)
$$\sigma_f^2 = \frac{\pi^2}{3} - 2 = 1.29.$$

Since α lies between 0 and π, it is clear from (34) and (35) that the response variance invariably exceeds the reinforcement variance.

In Table 1, σ_u^2 is given for various values of α. It is also instructive to

TABLE 1

Theoretical Variance of the Asymptotic Response Distribution for Various Values of α

α	U-theory	S-theory	I-theory
0	3.290	3.290	2.290
9°	3.239	3.221	2.294
18°	3.186	3.149	2.306
36°	3.082	3.002	2.354
90°	2.866	2.653	2.653
120°	2.876	2.628	2.876
150°	3.017	2.812	3.099
180°	3.290	3.290	3.290

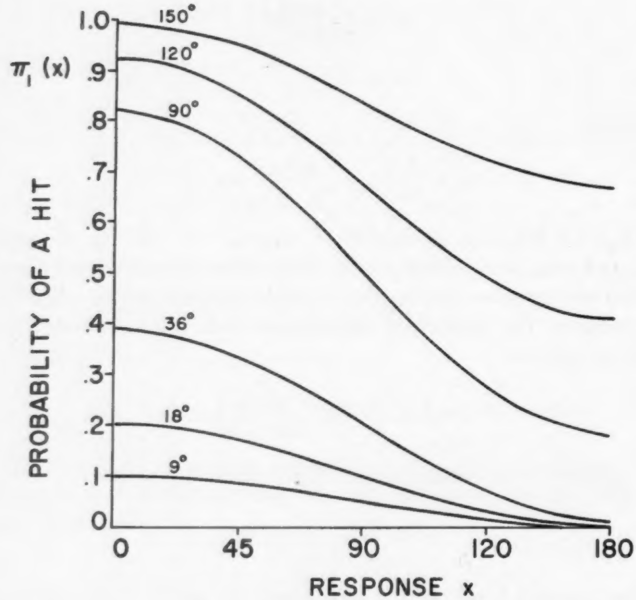

FIGURE 1

The probability of reinforcement for various values of the angle α. Due to their symmetry only half the curves are given. The reinforcement-function $f(y)$ which has been assumed is given in (22).

compare these σ^2 values with the plot of $\pi_1(x_n)$ vs. x_n for corresponding values of α given in Figure 1.

Symmetric Theory

In the symmetric theory, the previous assumption for the k_m distribution is replaced by

$$k_m(x, x_n) = k_h(x, x_n - \pi) \tag{36}$$

while the remaining assumptions of the uniform theory are applicable here as well. Under these assumptions (13) becomes

$$r(x) = \int k_h(x, x_n)\pi_1(x_n)r(x_n)\,dx_n + \int k_h(x, x_n - \pi)\pi_0(x_n)r(x_n)\,dx_n\,. \tag{37}$$

To compare the predictive character of the uniform and symmetric models we shall solve (37) for the asymptotic response distribution, $r(x)$, taking for k_h and $f(y)$ the functions assumed previously in (21) and (22). The method

of solution follows the same pattern outlined in solving (25) so that we shall omit the details here. From (37)

$$r(x) = \frac{1}{2\pi}\left(1 + \frac{2 \sin \alpha}{3\pi - 2\alpha} \cos x\right), \tag{38}$$

which implies a response variance of

$$\sigma^2_{\mathrm{sym}} = \frac{\pi^2}{3} - \frac{4 \sin \alpha}{3\pi - 2\alpha}. \tag{39}$$

The similarity of the density functions in (33) and (38) indicates that with regard to the asymptotic response distribution the uniform and symmetric models do not behave very differently. The two response distributions differ mainly in their variances, the variance associated with the uniform model being consistently larger. It will be noted in Table 1 that the response variance as a function of α has a minimum value in each theory, although the minimum variance occurs at different values of α. In the uniform theory the response distribution $r(x)$ has a minimum variance when $\alpha = 1.79$ radians (103°) while in the symmetric theory the minimum variance occurs when $\alpha = 1.91$ radians (110°).

Although both theories lead to similar asymptotic response distributions they do have some different characteristics which are brought out by considering their sequential predictions. Unfortunately, however, the sequential statistics are not easily obtained with these linear models since simple recursions do not arise.

Stimulus Sampling Theory

In the finite version of the stimulus sampling theory it is usually assumed that each element comprising the stimulus set is conditioned to at most one response at a given time. (An element may be in an unconditioned state, in which case the element is conditioned to no response.) The continuous version of the theory is obtained by relaxing this restriction and permitting an element to be conditioned to a range of responses. The conditioning state of an element is then represented by a density function $k(x, z)$, where z denotes the mean of the distribution and x a value of the response random variable. We shall further assume here, since we are mainly concerned with periodic continua, that the distribution implied by this density function is symmetric about the mean z. (In [2], $k(x, z)$ is called a *smearing* density.)

The sequence of events which takes place on a given trial is then conceptualized as follows. (See [2] for a more detailed statement.) A stimulus element is drawn by the subject from the set available on the given trial which, we shall assume in this paper, consists of exactly one element. In this special case, the state of the organism on each trial is given by a value of a single parameter, viz., the mean z; knowing the value of z, and, of course,

the smearing density $k(x, z)$, we may determine the probability of the response x falling within any specified interval on the response continuum. In practice, however, the mean z is not known to the experimenter, and it is more meaningful to discuss the density of z on trial n, denoted by $g_n(z)$, and to define the response density $r_n(x)$ in terms of $g_n(z)$; that is,

$$r_n(x) = \int k(x, z)g_n(z)\,dz. \tag{40}$$

It will be seen that it is more convenient to express the conditioning assumptions of the stimulus sampling theory in terms of recursions involving $g_n(z)$ rather than $r_n(x)$. Since (40) can be used to obtain $r_n(x)$ when $g_n(z)$ is known, this procedure will not result in any loss of generality.

Following the occurrence of the response x_n in the simple non-determinate case, four outcomes are possible. These outcomes are described by two dichotomous random variables, the reinforcement variable Y_n, and a variable denoted by F_n (or $F_{i,n}$) which specifies whether the reinforcing event is "effective." When $Y_n = 1$ and $F_n = 1$ (or given $F_{1,n}$), i.e., when the subject has been informed he is correct and this reinforcing event is effective, it is assumed that the mean of the k distribution shifts to the point of the response, that is, $z_{n+1} = x_n$. To obtain the recursion which indicates how the density of z changes after each trial, we shall need the joint density of the events z_{n+1} (or x_n), $F_{1,n}$, and $Y_{1,n}$. If θ denotes the probability of $F_n = 1$, then

$$j_{n+1}(z, F_{1,n}, Y_{1,n}) = \theta\pi_1(z)r_n(z). \tag{41}$$

With probability $1 - \theta$ the reinforcement is not effective ($F_n = 0$). Then the mean of the k distribution does not shift, but remains fixed for the subsequent trial, that is, $z_{n+1} = z_n$. The joint density associated with these events is

$$j_{n+1}(z, F_{0,n}, Y_{1,n}) = (1 - \theta)g_n(z) \int k(x, z)\pi_1(x)\,dx. \tag{42}$$

It is further assumed that when the subject is incorrect ($Y_n = 0$) and the reinforcement is not effective ($F_n = 0$) that $z_{n+1} = z_n$; thus

$$j_{n+1}(z, F_{0,n}, Y_{0,n}) = (1 - \theta)g_n(z) \int k(x, z)\pi_0(x)\,dx. \tag{43}$$

For the effect of the fourth possible outcome ($F_{1,n}$ and $Y_{0,n}$) on z_n, we shall consider three possible alternatives:

(I) $\quad z_{n+1} = z_n,$

(S) $\quad z_{n+1} = x_n + \gamma,$

(U) $\quad g(z_{n+1} \mid F_{1,n}, Y_{0,n}) = \dfrac{1}{2\pi}.$

Assumptions (S) and (U), respectively, lead, as will be seen, to stimulus sampling analogues of the symmetric model (when $\gamma = \pi$) and the uniform model discussed previously. Assumption (I) leads to a distinct theory we shall term the *Identity* (or I) *theory*. The joint densities corresponding to these three assumptions are as follows:

$$
\begin{array}{lll}
 & \text{(I)} & j_{n+1}(z, F_{1,n}, Y_{0,n}) = \theta g_n(z) \int \pi_0(x) k(x, z)\, dx; \\
(44) & \text{(S)} & j_{n+1}(z, F_{1,n}, Y_{0,n}) = \theta r_n(z - \gamma)\pi_0(z - \gamma); \\
 & \text{(U)} & j_{n+1}(z, F_{1,n}, Y_{0,n}) = \dfrac{\theta}{2\pi} \int \pi_0(x) r_n(x)\, dx.
\end{array}
$$

By combining (41), (42), (43), and one of the equations given in (44) a recursion for $g_n(z)$ may be obtained. We shall consider separately each of the three possibilities.

Identity Theory

Equations (42) and (43) may be combined and simplified immediately to give

$$
\begin{aligned}
(45) \qquad j_{n+1}(z, F_{0,n}) &= (1 - \theta) g_n(z) \int k(x, z)(\pi_0(x) + \pi_1(x))\, dx \\
&= (1 - \theta) g_n(z).
\end{aligned}
$$

Combining (45) with (41) and (44I) gives the recursion

$$
(46) \qquad g_{n+1}(z) = (1 - \theta) g_n(z) + \theta\left[\pi_1(z) r_n(z) + g_n(z) \int \pi_0(x) k(x, z)\, dx\right].
$$

Asymptotically this reduces to

$$
g(z) = \frac{\pi_1(z) r(z)}{1 - \int \pi_0(x) k(x, z)\, dx} = \left[\frac{\pi_1(z)}{\int \pi_1(x) k(x, z)\, dx}\right] r(z),
$$

which we abbreviate by introducing $H(z)$ as follows:

$$
(47) \qquad g(z) = H(z) r(z).
$$

From (47) and (40) the asymptotic expressions for $r(x)$ can be obtained; multiplying (47) by $k(x, z)$, integrating over z, and using (40) we have

$$
(48) \qquad r(x) = \int k(x, z) H(z) r(z)\, dz.
$$

If we take for $k(x, z)$ and $\pi_1(x)$ the functions in (21) and (23), respectively, (48) may be solved explicitly for $r(x)$. The result of this solution is

$$r(x) = \frac{1}{2\pi}\left(1 + \frac{\sin \alpha}{2\alpha} \cos x\right), \tag{49}$$

which has a variance of

$$\sigma_i^2 = \frac{\pi^2}{3} - \frac{\sin \alpha}{\alpha}. \tag{50}$$

Comparing (50) with the variance expressions of the linear uniform and symmetric theories given in (34) and (39) indicates at least one significant difference. The minimum variance in the identity theory occurs as α approaches zero, or, alternatively, the variance increases monotonically as α increases. In the uniform and symmetric theories, however, the variance has a minimum when α equals 103 and 110 degrees, respectively. (See Table 1 for comparative values of the variance.) Thus varying α would appear to be one way of discriminating between the identity theory and the two linear theories. (It should be noted that although it is possible to formulate the identity model within the linear theory context, simple closed-form expressions do not result for this model and for this reason it was not discussed previously.)

Uniform Theory

The appropriate recursion obtained by summing (44U) with (41) and (45) is

$$g_{n+1}(z) = \theta\pi_1(z)r_n(z) + \frac{\theta}{2\pi}\int \pi_0(x)r_n(x)\,dx + (1 - \theta)g_n(z), \tag{51}$$

which gives asymptotically

$$g(z) = \pi_1(z)r(z) + \frac{1}{2\pi}\int \pi_0(x)r(x)\,dx.$$

The asymptotic response density $r(x)$ is then equal to

$$r(x) = \int k(x, z)\pi_1(z)r(z)\,dz + \frac{1}{2\pi}\iint k(x, z)\pi_0(x')r(x')\,dz\,dx', \tag{52}$$

but due to the symmetry assumption on the k distribution $k(x, z) = k(z, x)$ so that the integration over z equals one. Thus (52) becomes

$$\begin{aligned} r(x) &= \int k(x, z)\pi_1(z)r(z)\,dz + \frac{1}{2\pi}\int \pi_0(x')r(x')\,dx' \\ &= \int k(x, z)\pi_1(z)r(z)\,dz + \frac{1}{2\pi} - \frac{1}{2\pi}\int \pi_1(x')r(x')\,dx' \end{aligned} \tag{53}$$

which clearly agrees with (25). Thus the two uniform theories lead to the same asymptotic response distribution.

Symmetric Theory

To obtain the recursion in $g_n(z)$ for the symmetric model, we let $\gamma = \pi$ in (44S) and combining (44S) with (41) and (45) gives

$$(54)\qquad g_{n+1}(z) = (1 - \theta)g_n(z) + \theta[\pi_1(z)r_n(z) + r_n(z - \pi)\pi_0(z - \pi)].$$

At asymptote

$$(55)\qquad g(z) = \pi_1(z)r(z) + \pi_0(z - \pi)r(z - \pi),$$

or in terms of $r(x)$

$$(56)\qquad r(x) = \int_{-\pi}^{\pi} k(x, z)\pi_1(z)r(z)\,dz + \int_{-\pi}^{\pi} k(x, z)\pi_0(z - \pi)r(z - \pi)\,dz.$$

Equation (56) is in fact precisely equivalent to the asymptotic response density of the linear-symmetric theory given in (37). To show this more explicitly we perform a change of variable in the second integral of (56). Let $(z - \pi) = z'$; the second integral then becomes

$$(57)\qquad \int_{-2\pi}^{0} k(x, z' + \pi)\pi_0(z')r(z')\,dz';$$

but since the functions $k(x, z' + \pi)$, $\pi_0(z')$, and $r(z')$ are periodic (with period equal to 2π), we may change the limits of integration back to $(-\pi, \pi)$. Thus identifying $k_h(x, x_n)$ with $k(x, z)$ the equivalence between (56) and (37) is explicitly realized. Thus the two symmetric models give rise to the same asymptotic response distribution. This is not to suggest, however, that the two theories are identical.

Sequential Characteristics

As a final topic we discuss the sequential statistic $r_{n+1}(x \mid Y_{0,n})$. First consider $g_{n+1}(z \mid Y_{0,n})$ defined by

$$(58)\qquad g_{n+1}(z \mid Y_{0,n}) = \frac{j_{n+1}(z, Y_{0,n})}{P(Y_{0,n})}.$$

The numerator of (58) for the identity theory is obtained by summing (43) and (44I); thus from (58)

$$(59)\qquad g_{n+1}(z \mid Y_{0,n})$$

$$= \frac{(1 - \theta)g_n(z)\int k(x, z)\pi_0(x)\,dx + \theta g_n(z)\int k(x, z)\pi_0(x)\,dx}{\int \pi_0(x)r_n(x)\,dx},$$

which is, of course, equal to

$$(60)\qquad g_{n+1}(z \mid Y_{0,n}) = \frac{g_n(z) \int k(x, z)\pi_0(x)\, dx}{\int \pi_0(x) r_n(x)\, dx},$$

indicating that this statistic is independent of θ. The asymptotic conditional response density

$$\lim_{n\to\infty} r_{n+1}(x \mid Y_{0,n})$$

is obtained directly from (60) by the usual procedure of first multiplying both sides by $k(x, z)$, substituting in (47), and then integrating over z:

$$(61)\qquad \lim_{n\to\infty} r_{n+1}(x \mid Y_{0,n}) = \frac{r(x) - \int k(x, z)\pi_1(z) r(z)\, dz}{\int \pi_0(x) r(x)\, dx}.$$

Using the functions for $r(x)$, $k(x, z)$, and $\pi_1(z)$ given previously in (49), (21), and (23), respectively, the asymptotic conditional response density is

$$(62)\qquad \lim_{n\to\infty} r_{n+1}(x \mid Y_{0,n}) = \frac{1}{2\pi}\left[1 + \left(\frac{(2\pi - 3\alpha)\sin\alpha}{4\alpha\pi - 4\alpha^2 - \sin^2\alpha}\right)\cos x\right].$$

For the uniform and symmetric theories, $g_{n+1}(z \mid Y_{0,n})$ is obtained by summing (43) with (44U) and (43) with (44S), respectively. In the former case

$$(63)\qquad g_{n+1}(z \mid Y_{0,n}) = \frac{(1 - \theta) g_n(z) \int k(x, z)\pi_0(x)\, dx + \frac{\theta}{2\pi} \int \pi_0(x) r_n(x)\, dx}{\int \pi_0(x) r_n(x)\, dx}$$

and, for $\theta = 1$,

$$(64)\qquad g_{n+1}(z \mid Y_{0,n}) = \frac{1}{2\pi},$$

indicating that

$$(65)\qquad r_{n+1}(x \mid Y_{0,n}) = \frac{1}{2\pi}.$$

In the symmetric theory,

$$g_{n+1}(z \mid Y_{0,n}) = \frac{(1-\theta)g_n(z)\int k(x, z)\pi_0(x)\,dx + \theta r_n(z-\gamma)\pi_0(z-\gamma)}{\int \pi_0(x)r_n(x)\,dx}, \tag{66}$$

which, unlike (60), depends on θ. If $\theta = 1$,

$$g_{n+1}(z \mid Y_{0,n}) = \frac{\pi_0(z-\gamma)r_n(z-\gamma)}{\int \pi_0(x)r_n(x)\,dx}, \tag{67}$$

giving for the response density

$$r_{n+1}(x \mid Y_{0,n}) = \frac{\int k(x, z+\gamma)r_n(z)\pi_0(z)\,dz}{\int \pi_0(x)r_n(x)\,dx}. \tag{68}$$

We remark that for $\theta = 1$, (65) and (68) also hold in the uniform and symmetric linear models, respectively.

It should be pointed out perhaps that (63) and (66) can be used to estimate θ for the symmetric and uniform theories so that definite predictions can be made for other sequential characteristics. Expressions such as $r_{n+1}(x \mid Y_{1,n})$ can easily be obtained in these theories and in all cases they show a dependence on θ.

The detailed application of the theoretical results obtained to appropriate experiments follows along exactly the same lines as the application in [3] of the results for determinate reinforcement.

Extensions

The ideas developed in this paper may be directly extended to other continua than the circumference of a circle. More importantly, the nondeterminate conditions of reinforcement may be varied in several directions without disturbing the main lines of the mathematical arguments leading to the mean asymptotic response distributions or some of the simpler sequential statistics. For instance, it is clear how to modify the theory of the basic target experiment described at the beginning in order to predict response behavior when the subject is shown the exact position of the target on those trials on which he misses. Another possibility of some interest is to show the subject the target's exact position on a certain proportion of the trials independently of the correctness or incorrectness of his response.

REFERENCES

[1] Suppes, P. A linear learning model for a continuum of responses. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univ. Press, 1959. Pp. 400–414.

[2] Suppes, P. Stimulus sampling theory for a continuum of responses. In K. J. Arrow, S. Karlin, and P. Suppes (Eds.), *Mathematical methods in the social sciences*. Stanford: Stanford Univ. Press, 1960. Pp. 348–365.

[3] Suppes, P. and Frankmann, R. W. Test of stimulus sampling theory for a continuum of responses with unimodal noncontingent determinate reinforcement. *J. exp. Psychol.*, 1961, 60, 122–132.

Manuscript received 11/7/60

Revised manuscript received 4/15/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

TWO LEARNING MODELS FOR RESPONSES MEASURED ON A CONTINUOUS SCALE*

NORMAN H. ANDERSON
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Two linear operator models are presented for a class of learning situations in which the response is on a numerical scale and the subject is given the magnitude of his error on some or all of the trials. Theoretical expressions are developed for sequential dependencies, mean learning curves, variances, and covariances, which permit a number of tests of goodness of fit.

The recent intensive development of stochastic learning models has been largely restricted to situations with discrete response classes, and there has been relatively little work on tasks with responses measured on a numerical scale. A number of articles have considered latency [e.g., 1, 4, 5, 6] but most of these models have been derivative from discrete response models rather than taking latency as the basic underlying variable. Anderson and Hovland [3] have applied a linear operator model to opinion formation, and Suppes [10] has also developed a linear model in which the operator acts directly on the numerical response measure. However, both of these models would seem to have somewhat limited potential applicability.

The purpose of this article is to discuss two models for situations involving a numerical response. Both are based on linear operators analogous to the stochastic models of Bush and Mosteller [5]. However, the present operators act directly on the numerical response, and the Bush-Mosteller development does not apply in this case.

It will be helpful to keep the following illustrative experimental situation in mind. The subject's task is to produce a line of some length. On each trial, he is told his error, i.e., how far he is from the correct length. It will be supposed that there are in general a number of "correct" lengths, each being chosen with a certain fixed probability on any trial. The models will be developed in the context of this situation. Possible applications of the models to other types of tasks are noted in the discussion section.

On each trial, a given subject will have some probability distribution over the response continuum. If the subject is told his error on any trial, there will result, presumably, a change in this probability distribution, the change being greater the greater the error. The two models considered here

*This work was supported by Grant G-12986 from the National Science Foundation. The author wishes to express his appreciation to J. H. Alexander for his cogent assistance.

assume that this change involves only the mean of the distribution. Specifically, the formal statements of the models are

Model I: $\mu_{n+1} = x_n - \alpha(x_n - d),$

Model II: $\mu_{n+1} = \mu_n - \beta(x_n - d),$

where μ_n is the response mean on trial n, x_n is the overt response on trial n, d is the length called correct on the given trial, and α, β are learning parameters. As is stated more precisely below, it is also assumed that any two probability distributions of response have the same shape and differ only in their means.

The models agree in assuming that the amount of change induced on any trial is a constant proportion of the error on that trial. They differ in what is changed. For Model II, the change occurs in the underlying mean, μ_n. The change is zero when either the error or the learning rate is zero; complete compensation occurs when $\beta = 1$. For Model I, the change acts on the observed response assuming a baseline of $\alpha = 0$ in which case the new mean is just the previous response, a sort of contiguity learning. When $\alpha = 1$, learning occurs in one trial. Although Model I may seem less plausible than Model II, it will be seen that they agree in many of their predictions.

Except in extreme cases, both models imply that the subject is continually influenced by the error information he receives. Consequently, even the asymptotic data obtained by using a single correct point may be used in testing the models.

Notation

Random variables are distinguished from their possible values by the use of asterisks. Thus, if x^* is any random variable, its (cumulative) distribution function, say $G(x)$, is by definition, $G(x) = \text{Prob}\ (x^* \leq x)$. If $G'(x)$ exists, this derivative is the frequency function or probability density function of the random variable, and is denoted by the corresponding lower case letter, $g(x) = G'(x)$. When it will cause no confusion, the asterisk will be omitted from the random variables. Thus, expected values will be written as $E(x)$ instead of $E(x^*)$, variances as σ_x^2 instead of $\sigma_{x^*}^2$, etc.

The symbol $dG(x)$ denotes the Lebesgue-Stieltjes differential. It is convenient to think of the differential as the amount of probability at x, $dG(x) = \text{Prob}\ (x^* = x)$. Although this verbal interpretation is heuristic, the use made of the differential here is correct.

If $G(x)$ is differentiable, then $dG(x) = g(x)\ dx$. The reader who is more familiar with frequency functions may prefer to interpret the various integrals by changing the differentials to this form. In terms of these notations, the expected value of x^* is

$$E(x) = \int x\ dG(x) = \int x\ g(x)\ dx.$$

The distribution of the response random variable on trial n is denoted by $R(x_n)$ so that $R(x_n) = \text{Prob}\,(x_n^* \leq x_n)$.

For a given subject, there will in general be a number of different possible reinforcement sequences leading up to trial n. Consequently, there will be a number of different possible response means for that subject on any trial. From this standpoint, the mean for a single subject is itself to be considered as a random variable. In order to eliminate subscripts in the derivations, the means on trials n, $n + 1$, and $n + 2$ will be denoted by μ^*, ν^*, and ϕ^*, with distributions $M(\mu)$, $M(\nu)$, and $M(\phi)$, respectively.

If $\mu^* = \mu$ for a given subject, then the actual response on trial n is governed by a probability distribution centered at μ. Let $G(x_n) = \text{Prob}\,(x_n^* \leq x_n)$, and let $F(x_n - \mu) = G(x_n)$. Then $F(x_n - \mu)$ is the distribution of the response random variable, $x_n^* - \mu$, conditional on the value of μ^*. This notation incorporates the assumption that the distributions of responses about any two means are identical except for their central points.

If

$$t^* = x_n^* - \mu,$$

then

$$E(t) = \int t\, dF(t) = \int (x_n - \mu)\, dF(x_n - \mu) = 0.$$

This relation, in various forms, will be employed extensively in the derivations.

The set of possible correct lengths in the experimental situation will be denoted by d_i, $i = 1, 2, \cdots, k$. It is assumed that each d_i occurs on any trial with some specified probability π_i. The mean value of the d_i is

$$\bar{d} = \sum \pi_i d_i .$$

The use of an i subscript on a function or operator will denote a conditional dependence on the occurrence of d_i.

Since many of the derivations use the same reasoning, a detailed development will be made only for the early results except when something new is involved. In particular, the derivations will be given for Model II and only listed for Model I. For one whole class of results, moreover, the expressions are formally identical for the two models. This will be indicated by suffixing "ab" to the equation numbers, thereby stating that a result valid for Model I may be obtained simply by substituting α for β throughout.

Sequential Dependencies

The results of this section come from consideration of the conditional distributions of means on trial $n + 1$, given that d_i was reinforced on trial n, and analogous higher order conditional distributions.

Let

$$M_i(\nu) = \text{Prob}\,(\nu^* \leq \nu \mid d_i \quad \text{on trial} \quad n),$$

where ν^* is the response random variable on trial $n + 1$. If $\mu^* = \mu$, and $x_n^* = d_i + (\mu - \nu)/\beta$, then the mean on trial $n + 1$ is ν according to the original statement of Model II. Hence, the probability that $\nu^* = \nu$, namely, $dM_i(\nu)$, is the probability that $\mu^* = \mu$, namely, $dM(\mu)$, times the probability that $x_n^* = d_i + (\mu - \nu)/\beta$, namely, $dF(d_i + (\mu - \nu)/\beta - \mu)$, integrated over all values of μ:

$$(1a) \qquad dM_i(\nu) = \int_\mu dF\left(\frac{\mu(1 - \beta) - \nu + \beta\, d_i}{\beta}\right) dM(\mu);$$

$$(1b) \qquad dM_i(\nu) = \int_\mu dF\left(\frac{\nu - (1 - \alpha)\mu - \alpha\, d_i}{1 - \alpha}\right) dM(\mu).$$

These two integral equations for the distributions of response means are basic to this section. To obtain the first result, let E_i be the expectation conditional on the occurrence of d_i on trial n. Then

$$E_i(\nu) = \int_\nu \nu\, dM_i(\nu) = \int_\nu \int_\mu \nu\, dF\left(\frac{\mu(1 - \beta) - \nu + \beta\, d_i}{\beta}\right) dM(\mu),$$

where the first equality is by definition and the second follows from (1a). To evaluate this expression, interchange the order of integration, and make the following substitution,

$$t = [\mu(1 - \beta) - \nu + \beta\, d_i]/\beta,$$

$$\nu = \mu(1 - \beta) + \beta\, d_i - \beta t.$$

Then

$$E_i(\nu) = \int_\mu \left\{\int_t [\mu(1 - \beta) + \beta\, d_i - \beta t]\, dF(t)\right\} dM(\mu)$$

$$= \int_\mu [\mu(1 - \beta) + \beta\, d_i]\, dM(\mu)$$

$$= (1 - \beta)E(\mu) + \beta\, d_i\ .$$

The second equality involves two steps which may need comment. The integral of $t\, dF(t)$ is 0 as noted in the previous section. Although t implicitly involves μ, the latter is to be considered constant in the first integration by definition of the conditional distribution F. An analogous derivation holds for Model I so that we obtain for the two models,

$$(2ab) \qquad E_i(\nu) = (1 - \beta)E(\mu) + \beta\, d_i\ .$$

The unconditional expected mean on trial $n + 1$, $E(\nu)$, is simply the sum of the conditional expectations weighted by π_i. Thus (2ab) implies

$$(3ab) \qquad E(\nu) = \sum \pi_i E_i(\nu) = (1 - \beta)E(\mu) + \beta\, \bar{d}.$$

Equations (3ab) are the difference equations for the mean learning curves as may be seen more clearly by changing notation for the moment to let $E(\mu) = \bar{\mu}_n$ and $E(\nu) = \bar{\mu}_{n+1}$. Rewriting (3ab) yields

$$\bar{\mu}_{n+1} = (1 - \beta)\bar{\mu}_n + \beta\, \bar{d};$$

solving this standard difference equation yields the mean learning curve

$$\bar{\mu}_n = \bar{d} - (\bar{d} - \bar{\mu}_1)(1 - \beta)^{n-1}, \tag{4ab}$$

where $\bar{\mu}_1$ is the mean response on the initial trial.

The most interesting implication of (4ab) is that for both models the asymptotic mean of the distribution of response means is the weighted average of the d_i, namely $\bar{d} = \sum \pi_i\, d_i$. This result is a matching theorem analogous to Estes' [7] result for categorical response situations.

It may be noted that if the learning rate depends on the value of the point of reinforcement, then the asymptote is $\overline{\beta\, d}/\bar{\beta}$. However, the case of unequal β will not be further considered here.

Higher order dependencies may be studied in a similar manner. Letting $M_{ij}(\phi)$ denote the distribution of means on trial $n + 2$, conditional on the occurrence of d_i on trial n and d_j on trial $n + 1$, reasoning similar to that above leads to the equation,

$$dM_{ij}(\phi) = \iint dF\left(\frac{\nu(1 - \beta) - \phi + \beta\, d_j}{\beta}\right) dF\left(\frac{\mu(1 - \beta) - \nu + \beta\, d_i}{\beta}\right) dM(\mu).$$

Multiplying through by ϕ and integrating over ϕ, ν, and μ yields

$$E_{ij}(\phi) = (1 - \beta)^2E(\mu) + \beta(1 - \beta)\, d_i + \beta\, d_j\, . \tag{5ab}$$

The integration is accomplished by using twice successively the t substitution employed above and noting that since F is a conditional distribution, ν is constant in the first differential, and μ is constant in the second.

Similarly, if ψ^* is the response mean on trial $n + 3$, given that d_i, d_j, and d_k occurred on trials n, $n + 1$, and $n + 2$, respectively, it can readily be shown that

$$E_{ijk}(\psi) = (1 - \beta)^3E(\mu) + \beta(1 - \beta)^2\, d_i + \beta(1 - \beta)\, d_j + \beta\, d_k\, . \tag{6ab}$$

Similar procedures yield the second moments. Note first that the unconditional distribution of means on trial $n + 1$ is simply the weighted sum of the several conditional distributions. Thus

$$dM(\nu) = \sum \pi_i\, dM_i(\nu).$$

The second raw moment $E(\nu^2)$ is then obtained by integrating ν^2 over this unconditional distribution. Use of the t substitution thus yields

$$E(\nu^2) = \sum \pi_i[(1 - \beta)^2E(\mu^2) + \beta^2E(t^2) + \beta^2\, d_i^2 + 2\beta(1 - \beta)\, d_iE(\mu)].$$

This expression may be simplified using the following relations:

$$E(t^2) = \sigma_F^2 ,$$

$$E(\mu^2) = \sigma_{M,n}^2 + E^2(\mu),$$

$$E(\nu^2) = \sigma_{M,n+1}^2 + E^2(\nu),$$

$$E^2(\nu) = [(1 - \beta)E(\mu) + \beta \bar{d}]^2,$$

$$\sigma_d^2 = \sum \pi_i d_i^2 - \bar{d}^2,$$

where σ_F^2 is the variance of the distribution of responses around a given mean, $\sigma_{M,n}^2$ is the variance of the distribution of means on trial n, and σ_d^2 is the variance of the d_i. After simplification,

$$\text{(7b)} \qquad \sigma_{M,n+1}^2 = (1 - \beta)^2\sigma_{M,n}^2 + \beta^2\sigma_F^2 + \beta^2\sigma_d^2 ;$$

$$\text{(7a)} \qquad \sigma_{M,n+1}^2 = (1 - \alpha)^2\sigma_{M,n}^2 + (1 - \alpha)^2\sigma_F^2 + \alpha^2\sigma_d^2 .$$

This pair of difference equations is the first instance in which the models lead to different results.

We are here primarily interested in the asymptotic variances, denoted by omitting the trial subscript. Setting $\sigma_{M,n+1}^2 = \sigma_{M,n}^2 = \sigma_M^2$ in (7b) and (7a) yields

$$\text{(8b)} \qquad \sigma_M^2 = \frac{\beta}{2 - \beta} (\sigma_F^2 + \sigma_d^2);$$

$$\text{(8a)} \qquad \sigma_M^2 = \frac{(1 - \alpha)^2\sigma_F^2 + \alpha^2\sigma_d^2}{1 - (1 - \alpha)^2}.$$

Response Distributions

The preceding section has been concerned with the distributions of means. The present section serves to relate the previous results to the observed response.

Consider the distribution $R(x_n)$ of the observed response on trial n. Now Prob $(x_n^* = x_n)$ is equal to Prob $(x_n^* = x_n \mid \mu^* = \mu)$ times Prob $(\mu^* = \mu)$, integrated over all values of μ. The integral equation for $R(x_n)$ is obtained by substituting the appropriate differentials to yield

$$dR(x_n) = \int dF(x_n - \mu)\, dM(\mu).$$

The expected value of the response on trial n is then

$$
\begin{aligned}
E(x_n) &= \int x_n \, dR(x_n) \\
&= \iint x_n \, dF(x_n - \mu) \, dM(\mu) \\
&= \iint [(x_n - \mu) + \mu] \, dF(x_n - \mu) \, dM(\mu) \\
&= E(\mu).
\end{aligned}
$$

Similar arguments hold for the conditional and unconditional expectations of the response on trial $n + 1$ so that

(9ab) $$E(x_n) = E(\mu),$$

(10ab) $$E_i(x_{n+1}) = E_i(\nu),$$

(11ab) $$E(x_{n+1}) = E(\nu).$$

These three equations state simply that the expected value of the observed response equals the expected value of the underlying means, as would be anticipated, of course. Consequently the various expected values of means in the previous section may be replaced by the appropriate data statistics for parameter estimation and tests of goodness of fit.

The second moment of x_n^* is also of interest.

$$
\begin{aligned}
E(x_n^2) &= \int x_n^2 \, dR(x_n) \\
&= \iint x_n^2 \, dF(x_n - \mu) \, dM(\mu) \\
&= \iint [(x_n - \mu)^2 + 2\mu(x_n - \mu) + \mu^2] \, dF(x_n - \mu) \, dM(\mu) \\
&= \sigma_F^2 + E\,(\mu^2).
\end{aligned}
$$

Now

$$E(\mu^2) = \sigma_{M,n}^2 + E^2(\mu), \quad \text{and} \quad E(x_n^2) = \sigma_{x_n}^2 + E^2(x_n),$$

where $\sigma_{x_n}^2$ is the variance of the distribution of overt responses on trial n. By (9ab), $E^2(x_n) = E^2(\mu)$ so that

(12ab) $$\sigma_{x_n}^2 = \sigma_{M,n}^2 + \sigma_F^2 .$$

This equation is essentially an analysis of variance of the observed response into the variance of the response means plus the variance of the response around a given mean. It may be noted that (12ab) is valid regardless of the form of operator assumed in the models.

Covariance Relations

In this section the expressions for the covariance of the observed response on trials n and $n + k$ are developed. The derivations are most convenient if frequency functions rather than distribution functions are used. Accordingly, it will be assumed that all distributions are differentiable. The symbol g_i will stand for any frequency function of the indicated arguments, conditional on the occurrence of d_i on trial n.

To find the covariance of x_n^* , x_{n+1}^* , given d_i on trial n, consider the joint frequency function of x_{n+1}^* , x_n^* , μ^*, conditional on the occurrence of d_i on trial n. Making use of the well-known expansion of a joint frequency function in a product of conditional frequency functions,

$$g_i(x_{n+1}, x_n, \mu) = g_i(x_{n+1} \mid x_n, \mu) g_i(x_n \mid \mu) g_i(\mu).$$

Of course,

$$g_i(\mu) = M'(\mu) = m(\mu), \quad \text{and} \quad g_i(x_n \mid \mu) = F'(x_n - \mu) = f(x_n - \mu).$$

If x_n and μ are known, then $\nu = \mu - \beta(x_n - d_i)$. Hence

$$g_i(x_{n+1} \mid x_n, \mu) = f[x_{n+1} - \mu + \beta(x_n - d_i)].$$

The conditional expectation of $x_n^* x_{n+1}^*$ is thus

$$E_i(x_n x_{n+1}) = \iiint x_n x_{n+1} f[x_{n+1} - \mu + \beta(x_n - d_i)] f(x_n - \mu)\, m(\mu)\, dx_{n+1}\, dx_n\, d\mu.$$

The evaluation is accomplished by integrating successively over x_{n+1} , x_n , and μ, after rewriting $x_n x_{n+1}$ in the following form,

$$x_n x_{n+1} = x_n(x_{n+1} - \mu + \beta x_n - \beta\, d_i) - \beta(x_n - \mu)^2$$
$$+ (x_n - \mu)(\mu + \beta\, d_i - 2\beta\mu) + (1 - \beta)\mu^2 + \beta\, d_i \mu.$$

The integrals of the first and third terms on the right vanish and there results

$$E_i(x_n x_{n+1}) = -\beta\sigma_F^2 + (1 - \beta)E(\mu^2) + \beta\, d_i E(\mu).$$

This expression may be simplified using the fact that the expected value of a product of two random variables equals their covariance plus the product of their expected values, and applying (2b) and (10b) to $E_i(x_n x_{n+1})$ to yield the conditional covariance

(13b) $$\mathrm{Cov}_i\,(x_n , x_{n+1}) = (1 - \beta)\sigma_{M.n}^2 - \beta\sigma_F^2 ,$$

(13a) $$\mathrm{Cov}_i\,(x_n , x_{n+1}) = (1 - \alpha)(\sigma_{M.n}^2 + \sigma_F^2).$$

The unconditional covariances have the identical form as the conditional covariances as can be shown fairly easily.

It is also possible to obtain expressions for the covariance of x_n^*, x_{n+k}^*, conditional on any given sequence of the d_i on trials $n, n+1, \cdots, n+k-1$. The derivations are somewhat tedious but the results are

$$\text{(14ab)} \qquad \mathrm{Cov}_c\,(x_n\,,\,x_{n+k}) = (1-\beta)^{k-1}\,\mathrm{Cov}_c\,(x_n\,,\,x_{n+1}),$$

where the subscript c denotes the conditional feature.

Goodness of Fit

Without going into detail on questions of parameter estimation, it is appropriate to indicate some of the straightforward ways in which the models may be tested. It will be assumed (i) that such tests are based on asymptotic data and (ii) that the estimation is done for each subject separately. These two restrictions are easily met in the illustrative length production task although they would be more difficult to fulfill in other situations. It would, of course, be possible to apply the above results to the early learning data and/or pool over subjects, but either procedure has certain disadvantages. The early trials are not unlikely to involve adaptation effects not allowed for in the present form of the models. Pooling before estimating may introduce bias and will increase variability because of individual differences in parameters. The use of steady state data from each subject separately tends to avoid both problems [2]. Moreover, because of the probabilistic nature of the sequence of reinforcements, the asymptotic data will yield information even with a single reinforced point.

The present models have the ergodic property [2, 5]. As a consequence, the trial average of any statistic will, as the number of trials becomes large, approach the mean value of that statistic over the distribution on any given trial. The expressions derived above for a single trial may thus be replaced for estimation purposes by the corresponding trial average for a single subject. For instance, $E(\mu)$ is estimated by the mean response, $E_i(\nu)$ by the mean response averaged over those trials following d_i occurrences, σ_x^2 by the variance of the set of responses, etc.

The numerical matching theorem (4ab) requires no parameter estimation but only a direct test of the subject's asymptotic response level against the a priori value $\bar{d}$. If matching does occur, than $E(\mu)$ in (2ab) may be replaced by the theoretical value $\bar{d}$. The conditional means of (2ab) then yield k linear equations in the single unknown learning rate parameter.

The higher order dependencies, being nonlinear, are most easily used as checks on goodness of fit. A relatively simple way to employ them is by taking differences as suggested by the results of [2] for the discrete response case. Replacing ϕ and ψ by x in order to emphasize that we are concerned with the asymptotic response, (5ab) and (6ab) yield

$$\text{(15ab)} \qquad E_{ik}(x) - E_{jk}(x) = \beta(1-\beta)(d_i - d_j),$$

(16ab) $$E_{ikm}(x) - E_{jkm}(x) = \beta(1 - \beta)^2(d_i - d_j),$$

for all values of i, j, k, and m.

The equations for the higher moments also yield information on the adequacy of the models. Thus (13a) and (13b) state that all asymptotic first-order conditional covariances have the same value within each model. Analogous considerations hold for the higher order covariances of (14ab). Since the models differ in their expressions for the covariances, these expressions thus give a basis for assessing the relative adequacy of the models.

These variance and covariance relations may also be combined to yield the asymptotic autocorrelations which bring out the difference between the models more clearly. For Model II, (8b), (12b), (13b), (14b), and the fact that the correlation is the covariance divided by the root product of the variances give

(17b) $$\rho(x_n , x_{n+k}) = -\tfrac{1}{2}\beta(1 - \beta)^{k-1}[1 - \sigma_d^2/\sigma_x^2].$$

Similarly, from (12a), (13a), and (14a), Model I implies

(17a) $$\rho(x_n , x_{n+k}) = (1 - \alpha)^k.$$

The autocorrelations thus depend only on the learning rate for Model I. For Model II, however, the autocorrelations depend also on the variance σ_d^2 of the points of reinforcement, as may be seen by applying (8b) and (12b) to get

$$\sigma_x^2 = (2\sigma_P^2 + \beta\sigma_d^2)/(2 - \beta),$$

and substituting this expression in (17b). Manipulation of σ_d^2 should thus affect the autocorrelations for Model II but not for Model I. This result thus affords a direct experimental contrast between the two models.

The case of $k = 1$ deserves special comment. The sequential dependencies give no information when there is only one reinforced point, but the autocorrelation results may still be usefully applied. With a single reinforced point, $\sigma_d^2 = 0$. Hence, on the plausible assumption that the learning rates lie between 0 and 1, (17a) and (17b) show that the autocorrelations are all positive for Model I, and all negative for Model II.

Discussion

It is of some interest to compare the two models developed here with previous results. The Anderson-Hovland [3] model may, in present notation, be written as

$$\mu_{n+1} = \mu_n - \gamma(\mu_n - d).$$

Application of the techniques used above shows that (2)–(6) hold exactly for this model when β is replaced by γ. The expressions for the variances and

covariances are fairly similar to those for Model I. Indeed, (7a), (8a), (13a), and (14a) are true with α replaced by γ, and the term in σ_F^2 eliminated. These last results lead directly to the autocorrelations; from these it is seen that, when there is but a single reinforced point, the asymptotic autocorrelations are all zero. This implication thus gives a ready test of the Anderson-Hovland model against Models I and II.

Suppes' formulation [10] has a somewhat different structure from the models considered here, so that a complete comparison is difficult. However, this formulation apparently can be extended to the case of a single reinforced point and would then imply that the mean of the asymptotic response distribution for a given subject is constant over trials. Consequently, this formulation would also imply that the asymptotic autocorrelations are all zero in contrast to Models I and II.

When the experimenter's choice of the point to be reinforced is independent of the subject's response, both the Anderson-Hovland and the Suppes models assume that the change in response on any trial is independent of the actual response made. It thus seems unlikely that either model would be applicable to tasks similar to the length production task even though they might be applicable to experiments on opinion and impression formation.

Experimental situations with more than one reinforced point provide the simplest comparative tests of Models I and II. Moreover, the efficiency of the tests is increased by increasing the spread of the reinforcement points. In principle, therefore, the use of several widely separated d_i would seem desirable. In practice, however, the use of widely separated d_i may well reduce the task to a guessing game to which, judging from current experiments on probabilistic multiple-choice learning tasks, the present linear models would not be expected to apply. The use of a numerical response measure would probably prove interesting in such a situation, but some caution in choosing the reinforced points is indicated if a strict test of the present models is desired.

A seeming limitation of the models is the assumption that the values of the reinforced points are specified by the experimenter. It would seem possible that constant errors may be involved, so that with a single point of reinforcement, for example, the asymptotic mean would be different from that point. More generally, the limit point [5] of the operator for a given point of reinforcement may be somewhat different from that point. However, the models may be applied directly to this case by simply reinterpreting the d_i as empirical rather than nominal values. The derivations remain identical and the only change is the necessity for estimating the d_i from the data.

It should be noted that treating the d_i in this way suggests the possibility of applying the models to considerably different situations from the

length production task. One such possible application would be to speed in a straight runway as in the work of Logan [8]. In this example, the limit points of the operators could probably not be specified on a priori grounds but would require empirical determination. There would also be a greater interest in the initial behavior in contrast to the length production task.

There are a number of possible extensions of the models to cover more complicated reinforcement schedules such as those discussed by Estes [7], as well as the continuous reinforcement functions considered by Suppes [10]. The present techniques would appear adequate to handle those cases in which the reinforced point is not contingent on the response, but the contingent case presents more difficulty. These extensions are not considered here but there are two minor schedule variations which deserve comment.

Although it is not feasible to use a strictly noncontingent random schedule in which the amount of "error" is chosen independently of the response, it is possible to insert an occasional test trial of this type. On such a test trial, the pre-chosen error would be formally represented by $x_n - d$ in the models, but would in fact be independent of the actual response and of the underlying mean. By (9ab) and (10ab), the initial statement of the models could then be applied directly to the response on the test trial and the following trial to estimate the learning parameter. If the models are correct, the parameter estimates obtained from these test trials will agree with the estimates from the regular trials. However, if the models are inadequate, such test trials may be particularly useful in locating the source of inadequacy. It should also be noted that test trials of this type would probably be especially valuable when only a single point of reinforcement is used.

The second variation concerns the use of nonreinforced trials on which the subject is given no error information. One might expect that such trials would effect no change in the underlying mean. If so, a number of possibilities arise which can be illustrated by supposing that every fifth trial only is reinforced. All four nonreinforced trials following a reinforced trial would then provide duplicate information which would be especially valuable if the pre-asymptotic phase were being studied. The variance computed from the data of the nonreinforced trials would also provide a direct estimate of σ_F^2 and, with sufficient data, one could obtain a frequency histogram which estimates the entire conditional response distribution, $F(x - \mu)$. The correlations over the nonreinforced trials would also be relevant to the question of possible nonindependence of successive responses [11], or perseverative effect [9], not taken into account in the present formulation.

REFERENCES

[1] Anderson, N. H. Temporal properties of response evocation. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univ. Press, 1959. Pp. 125–134.

[2] Anderson, N. H. An analysis of sequential dependencies. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univ. Press, 1959. Pp. 248–264.

[3] Anderson, N. H. and Hovland, C. I. The representation of order effects in communication research. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven: Yale Univ. Press, 1957. Pp. 158–169.

[4] Audley, R. J. The inclusion of response times within a stochastic description of the learning behavior of individual subjects. *Psychometrika*, 1958, **23**, 25–31.

[5] Bush, R. R. and Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

[6] Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

[7] Estes, W. K. Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika*, 1957, **22**, 113–132.

[8] Logan, F. A. *Incentive*. New Haven: Yale Univ. Press, 1960.

[9] Sternberg, S. H. A path-dependent linear model. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univ. Press, 1959. Pp.308–339.

[10] Suppes, P. A linear model for a continuum of responses. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univ. Press, 1959. Pp. 400–414.

[11] Verplanck, W. S., Collier, G. H., and Cotton, J. W. Non-independence of successive responses in measurements of visual threshold. *J. exp. Psychol.*, 1952, **44**, 273–282.

Manuscript received 10/16/60

Revised manuscript received 4/4/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

AN EMPIRICAL STUDY OF THE FACTOR ANALYSIS STABILITY HYPOTHESIS*

HAROLD P. BECHTOLDT
UNIVERSITY OF IOWA

Note is taken of four related sources of confusion as to the usefulness of Thurstone's factor analysis model and of their resolutions. One resolution uses Tucker's distinction between exploratory and confirmatory analyses. Eight analyses of two sets of data demonstrate the procedures and results of a confirmatory study with statistical tests of some, but not all, relevant hypotheses in an investigation of the stability (invariance) hypothesis. The empirical results provide estimates, as substitutes for unavailable sampling formulations, of effects of variation in diagonal values, in method of factoring, and in samples of cases. Implications of these results are discussed.

It has been said that a work of art should provoke favorable or unfavorable reactions, and that a scientific theory should lead to further empirical and theoretical work. Factor analysis, which has been called both an art and a scientific approach to the study of individual differences, certainly has evoked strong emotional responses as well as extensive empirical and theoretical studies. However, neither reaction has succeeded in clarifying either the role of factor analysis or the appraisal of its usefulness as a research technique. One might say confusion, if not chaos, is the norm in this field.

In the development of a science of psychology, confusion about the usefulness of a set of procedures such as those of factor analysis should be a matter of great concern. This paper takes the position that the present confusion stems, in part, from disagreements as to definitions of terms or concepts, and, in part, from failures to make certain analytical distinctions. A recent symposium on the "Future of factor analysis" [42] exemplifies several of these semantic and analytical confusions. The objectives of the present paper are threefold: (i) to call attention to these sources of confusion, to some of their implications, and to procedures for resolving the confusion; (ii) to demonstrate a type of factor analysis in which the usefulness of some hypotheses related to the stability or invariance of factor analysis data can be evaluated; and (iii) to provide, for several factoring procedures, empirical estimates of the sampling variation in objectively determined oblique simple

*The computational costs of this study were defrayed, in part, by a research small grant M-1922 from the National Institute of Health, and, in part, by support under project 176-0002 by the University of Iowa Computing Center, Dr. J. P. Dolch, Director. The assistance of Dr. Kern Dickman and Mr. Leonard Wevrick of the University of Illinois and of Mr. Norman Luther of the University of Iowa in handling the computing problems is gratefully acknowledged.

structure values based on data from two samples. Since available solutions to some statistical problems associated with factor analyses are impractical, such empirical estimates are useful but limited substitutes for the desired analytical formulations.

Sources of Confusion

One of the most insidious and ubiquitous sources of confusion is the ambiguity of the term *factor analysis*. That different models are classed under this single term is well known, but discussions of the objectives, techniques, and results obtained often do not make clear the specific model under consideration. For example, the principal components model must be distinguished from factor analysis models. The several factor analysis models in turn are differentiated on the basis of the postulation of a general factor, the acceptance of the necessity for rotation, and the criteria for the final solution (orthogonal vs oblique axes, simple structure, etc.). Statements appropriate to one model often do not apply to another even though certain transformations from one model to another may exist.

The present discussion deals with the model as formulated by Thurstone [52, 54, 55] and extended or modified by the work of Anderson and Rubin [2], Bargmann [3, 4], Guttman [29, 30], Howe [36], Koopmans and Reiersöl [38], Lawley [39], Rao [46], Reiersöl [48], and Tucker [62, 63]. These papers indicate the relevance to factor analysis theory of the problems concerned with (i) the existence of the model (solvability), (ii) the identification of the parameters (uniqueness), (iii) the determination of the number of factors (rank), (iv) the criteria for rotational transformations, and (v) the test of the hypothesis that the model fits a set of data. The acceptability of Thurstone's formulation involving oblique simple structure with common and unique factors requires consideration of solutions to these problems, several of which Thurstone explicitly recognized ([54]; [55], pp. v–xiv). Unfortunately, the extensions and clarifications of Thurstone's earlier work in the articles noted above have not been considered in recent articles critical of this factor analysis model [32, 33, 42, 58, 59, 60, 66, 67]. A minor but frustrating related source of confusion is the introduction of different names for the operations and concepts of Thurstone's model [60]. Furthermore, the implications of these papers have not been adequately considered by some proponents [17, 26] who use this model of factor analysis with test data. The result has been a proliferation of irrelevant and unacceptable arguments as to the usefulness of Thurstone's method as well as that of any other model of factor analysis.

A second source of confusion derives from failures to make or to maintain distinctions related to the objectives of an investigation. Such differences in purpose exist between the type of factor analysis which Tucker [63] calls *exploratory* as opposed to the type that Tucker calls *confirmatory*.

Basically his distinction depends upon the amount of information and of precision of knowledge in an area. The exploratory factor analysis, being the first, is used to generate hypotheses while a confirmatory factor analysis is designed subsequently to test these hypotheses. It is generally accepted as a principle of hypothesis testing that the same set of data cannot be used both for inventing or generating hypotheses and for evaluating the usefulness of the hypotheses. It is, of course, conceivable that an initial analysis could be sufficiently precise to permit the use of the term *confirmatory* and to permit the testing of certain hypotheses.

The purpose of an exploratory analysis, as stated by Thurstone, is "... to discover the principal dimensions or categories ... and to indicate the directions along which they may be studied by experimental laboratory methods" ([54], p. 189; [55]). The principal dimensions are discovered by the appearance of trans-situational response consistencies defined by the operations of factor analysis discussed in detail by Thurstone [55] and Tucker [61, 63], for example. This objective also can be expressed as the development of definitions of new composite variables and as the invention of hypotheses involving such variables [8]. In either type of activity the creative, artistic judgment of an investigator is as relevant in an exploratory factor analysis as in other creative endeavors; but the prior compulsions of the investigator for orthogonality or for a general factor, for example, will also be represented in his judgments and formulations [18]. The reader of the factor analysis literature should recognize that different compulsions lead to different results, i.e., to factors differently defined. Factors from two or more studies logically are not the same factors unless the defining operations including the reference tests and factoring procedures are the same. They may be similar factors or even parallel factors, provided definitions of such terms are specified, as Gulliksen does for parallel tests [27].

Since the initial formulation or invention of a variable or of a hypothesis cannot be useful by fiat, subsequent research to evaluate this usefulness is necessary. For this latter purpose, one may compare the empirical results using new samples from the same population with those obtained in the initial investigation; this procedure checks the stability or the invariance of the factor pattern [55]. Another stability question deals with the consistency of the empirical relations among modified or improved reference variables with those observed with the initial or unimproved variables, i.e., invariance under changes in the stimulus-response features of the task [6, 26]. Other hypotheses formulated in the initial exploratory study may deal with the number of significant factors and the location of specified zero factor loadings. For the investigation of such questions, subsequent confirmatory factor analyses on new samples of cases would be appropriate. Completely objective techniques for the conduct of such studies, if they are properly designed, are available together with some of the desired statistical

tests. The design of such confirmatory studies is not a matter of guesses or hunches, nor can just any available table of correlations be used because specific and often testable hypotheses are involved. The problems of designing such studies have been considered repeatedly by Thurstone [52, 53, 54, 55] and by Tucker [63]; necessary and sufficient conditions for existence and uniqueness of solutions are indicated by Anderson and Rubin [2].

A third source of confusion in the literature is associated with the types of hypotheses that can be evaluated by factor analysis procedures. Some aspects of this distinction have been noted by Eysenck [21] and by Peel [44]. Only a few specific hypotheses of the many considered in the factor literature can be appropriately investigated with the conventional factor analysis models. For many hypotheses, a distinction is made, or implied, in the statements of the hypotheses between a set of reference variables and/or a set of treatment conditions as the independent variables on the one hand and the experimental or dependent variables being studied on the other. This distinction is associated with differences in status for these two classes of variables. For factor analyses, Thurstone specifically rejects this distinction between independent and dependent variables ([54]; [55], p. 59); in fact, the accepted principle in the several factor analysis models is that all variables are to be treated as coordinate or equal. Thus, for hypotheses expressing one or more variables as functions of one or more other variables, the usual factor analysis model is inappropriate (unless the modifications noted below are made). Such hypotheses include those dealing with the effects upon factor scores of variations in age, kind of instruction, amount of practice, drugs, or genetic history, for example. Other nonfactorial hypotheses include those dealing with factors as sources of variance in scores on tasks not used in the definitions of the factors. Both the distinction between independent and dependent variables and the introduction of approximations to part-whole correlations are points of issue in attempts to use a factor method for such hypotheses. In addition, the problem of communalities and the process of standardizing scores in computing correlations create further difficulties for between-group comparisons [47].

A fourth source of confusion arises from the description of factors as underlying causal variables which are not observable, which can only be inferred from the response consistencies, and which cannot be explicitly defined. This linguistic formulation involves a debated point in the philosophy of science, a point Bergmann [11, 12] calls the confusion between meaning and significance. A related argument is treated by Henrysson [34] in a discussion of explanatory factor analysis. The problem for factor analysis is that such unobservable variables cannot be directly studied as can other defined concepts. The factors cannot be investigated in the laboratory, for example, as suggested by Thurstone nor can the relations between factors and other variables be evaluated by nonfactorial methods, a procedure considered

important by Thurstone [54, 55] as well as by most experimental psychologists. The factors are treated as existential hypotheses or almost as reified entities [20]. Brodbeck has made several pertinent points regarding this manner of speaking [13, 14]. And such writers as Anderson and Rubin [2], Koopmans and Reiersöl [38], and Rao [46] also have noted some of the difficulties associated with the unobservable characteristic of factors.

The fourth source of confusion can readily be resolved by using the results of an exploratory factor analysis and possibly of one or more confirmatory analyses to provide explicit objective definitions of the factors. These definitions will specify a factor as a definite function of observations on one or more designated reference variables. Such definitions are consistent with the existence of such factored tests as Thurstone's PMA battery [57] or of such sets of factor reference tests as the ETS Kit [24]. This resolution is consistent with Thurstone's statement of the objective of factor analysis; it also has several important linguistic and empirical implications. For example, the identification of factors as the same factor is not a problem [10] nor are procedures for defining a factor space common to two test batteries [64]. The third source of confusion can then be resolved by using explicitly defined factors as predictors in combination with the separation of independent and dependent variables in the analysis. With these two modifications of the factor analysis model, the several factor techniques can be shown to be ways to compute beta weights in the linear regression model. A convenient computing procedure uses the operations of the multiple-group method to project the dependent variables onto the space of the independent or factor variables. This relation follows directly from the early work of Holzinger and Harman [35] and Young and Householder [68]. In addition, explicitly defined factors can be used in nonfactorial experiments either as independent variables or as dependent variables. When a factor is explicitly defined without restricting the sample means and variances, the scores on the defined factors can be used as any distribution of test scores is used. The usefulness both of hypotheses involving factors and of proposed definitions of a factor then can be evaluated by the procedures regularly used for other hypotheses and other concepts.

Confirmatory Factor Analysis

The present study provides a demonstration of a confirmatory factor analysis conducted with a set of objective procedures in an investigation of the invariance of a simple structure solution, i.e., of the stability hypothesis. A series of questions related to the testable hypotheses of a confirmatory factor analysis are investigated; the relevance of these questions has been emphasized by Maxwell [41]. Answers to the questions were obtained from data on two samples of cases for a set of seventeen reference variables hypothesized on the basis of previous factor studies to be associated with a

given number (six) of factors and with a specified set of zero and nonzero factor loadings. The following five questions are considered.

1. Is the hypothesis of two independent random samples from a single multivariate normal population tenable with reference to two sets of means and two variance-covariance matrices?

2. If the first hypothesis is tenable, can the set of 17 variables for each sample be considered as demonstrating some significant amount of dependency as defined by Bargmann ([4], pp. 43–68) (i.e., the rejection of the hypothesis of independence)?

3. If the first two hypotheses are tenable, does the degree of dependence (number of factors), as defined by the maximum likelihood or canonical correlation procedures for each sample, correspond to a hypothesized value—namely six?

4. If the first three hypotheses are tenable, does the factor pattern of zero and nonzero loadings for each sample as defined by an oblimax analytical rotation correspond to the hypothesized factor pattern for the results obtained from three factoring methods applied to the data, including one or more of four sets of estimated communalities?

5. Are the results of the simpler graphical (judgmental) rotational methods and of the multiple-group methods without rotation consonant with those obtained from other methods of analysis?

Data

The data are a portion of those originally collected by Thurstone and Thurstone ([56], ch. 3) for an analysis involving the hypothesis of seven primary mental abilities. Statistical tests of the relevant hypotheses were not then available. On the basis of the earlier analysis, the definition of the *Perceptual Speed* factor was judged by Thurstone to be inadequate, and it was therefore dropped from the present study of sampling effects. In addition, one of the variables for the *Memory* factor, the Figure Recognition test, was eliminated as being an unacceptable defining variable for the factor M. The 17 remaining tests were then considered as defining six primary mental abilities (PMA's) by six isolated constellations such that variations in the rotated factor loadings would provide a useful estimate of the sampling fluctuations for the statistics under investigation. Two samples of cases ($N = 212$ and $N = 213$) were formed by assigning each of 425 cases alternately to one or the other of two groups after the cases were thoroughly randomized. The original data in Tables 1, 2, and 9 were computed by Dorothy Case Bechtoldt in an unpublished study under the direction of L. L. Thurstone and L. R Tucker.

The list of 17 variables along with the means and standard deviations for these two samples as well as the hypothesized nonzero factor loadings for each variable are presented in Table 1. The location of each nonzero

TABLE 1

Seventeen Variables With Sample Means and Standard Deviations

Code No.	Name of Variable	Sample I (N=212) Mean	S.D.	Sample II (N=213) Mean	S.D.
1	First Names (M)	9.44	4.507	9.80	4.554
2	Word-Number (M)	4.77	3.602	5.44	3.626
3	Sentences (V)	13.42	4.730	13.75	4.651
4	Vocabulary (V)	27.03	10.317	26.71	10.797
5	Completion (V)	31.97	10.795	31.89	10.581
6	First Letters (W)	36.65	9.778	36.18	11.152
7	Four Letter Words (W)	11.08	4.655	10.85	5.312
8	Suffixes (W)	9.07	4.106	8.46	4.513
9	Flags (S)	25.08	12.427	24.44	11.256
10	Figures (S)	22.70	12.798	22.01	11.451
11	Cards (S)	26.45	13.215	24.85	11.523
12	Addition (N)	16.39	6.991	15.92	7.079
13	Multiplication (N)	32.26	13.430	33.32	12.501
14	Three-Higher (N)	27.21	8.740	25.93	9.840
15	Letter Series (R)	12.40	5.725	12.46	5.718
16	Pedigrees (R)	16.10	7.678	16.45	7.651
17	Letter Grouping (R)	13.32	4.171	13.35	3.879

value is designated by the letter in parentheses. The time limits and scoring formulas are given in Thurstone and Thurstone ([55], p. 28). Product moment correlations among the 17 variables were computed separately for the two samples as shown in Table 2.

Results and Discussion

The first question of interest has to do with the comparability of the means, variances, and covariances for these 17 variables in the two samples. The hypothesis of equal variance-covariance matrices was tested by the procedures given by Anderson ([1], ch. 10) and by Federer [22] and reviewed by Maxwell [41]. The determinant test indicated that the hypothesis of equal variance-covariance matrices was tenable ($\phi = -2 \ln \lambda = 148.697$ for 153 d.f., $p > .05$). The equality of the two sets of means for the 17 variables was evaluated by Hotelling's T^2 statistic ([1], ch. 5). The hypothesis of equal means for the two samples (but not the equality of the means within a sample) was tenable ($F = 1.129$ for 17 and 407 d.f., $p > .05$). Together these two tests indicated that the hypothesis of independent random sampling

TABLE 2

Product Moment Intercorrelations*

Code No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1		472	290	401	299	234	254	296	86	61	52	246	274	250	332	313	297
2	482		189	220	232	209	246	193	44	78	157	151	146	60	238	213	170
3	299	275		833	761	402	275	374	103	19	-77	332	297	352	536	567	468
4	331	303	828		772	446	358	473	109	45	105	335	352	384	507	514	404
5	266	273	776	779		394	275	426	342	227	294	329	254	438	490	512	430
6	335	273	439	493	460		627	516	176	104	95	355	365	354	404	365	375
7	342	199	432	464	425	674		480	161	138	49	354	327	318	330	275	317
8	333	290	447	489	443	590	541		79	7	12	288	284	280	327	323	285
9	124	169	117	121	193	178	223	118		672	606	286	189	379	289	277	287
10	32	85	51	77	180	81	192	7	593		728	164	49	236	160	165	181
11	77	193	151	146	174	158	239	114	651	684		171	32	251	200	208	207
12	151	287	268	312	263	241	180	181	208	109	210		651	517	439	320	399
13	259	258	319	344	291	338	295	234	179	144	195	661		546	435	293	452
14	279	223	359	356	342	290	344	298	362	273	331	536	548		512	442	456
15	307	260	447	432	401	381	402	288	252	203	257	361	379	440		671	622
16	447	293	541	537	534	350	367	320	85	129	151	206	298	438	555		538
17	274	216	380	358	359	424	446	325	270	203	293	311	329	410	598	452	

* The data for sample I (N=212) are shown above the principal diagonal and those for sample II (N=213) below the diagonal. Correlations are multiplied by 1000.

from a single multivariate normal distribution was reasonable. The distributions of scores for the factor M tests, variables 1 and 2, however, were somewhat positively skewed. Incidentally, only one of the pairs of sample variances (considered separately for each variable) is significantly different (for variable 11, $F = 1.32$, $.02 < p < .10$).

Since the hypothesis considered by factor analysis is that the 17 variables within each set are not independent, i.e., that one or more degrees of dependence are indicated [4], the hypothesis of independence (the second question) was investigated using the determinant test as given by Anderson ([1], ch. 9) and Bargmann [4] for both samples. The hypothesis of independence was rejected ($\theta = -m \ln V$; $\theta = 1890.303$ and 1857.811 with 136 d.f. for samples I and II, respectively; $p < .001$). The values of the determinants V of the correlation matrices were 8.6658×10^{-5} and 11.8518×10^{-5} for samples I and II, respectively. These results indicated that a factor analysis is justified for each set of data.

Since, for this investigation, the first two hypotheses were tenable, the hypothesized rank of six was then investigated. The canonical correlation (maximum likelihood) approach of Rao [46] was used as a test on the rank,

TABLE 3

Communality Estimates*

Method:	Multiple R squared (inverse)		Centroid high r (adjusted)		Centroid mult. R (15 cycles)		Centroid unity (20 cycles)		Max. like. mult. R (Rao)		Prin. axes unity		Clusters unity	
Sample:	I	II	I	II	I	II	I	II	I	II	I	II	I	II
Code No.														
1	370	390	454	474	382	789	452	606	396	731	712	759	747	751
2	312	323	499	486	762	328	606	417	681	367	794	780	747	751
3	728	747	841	807	838	809	832	805	835	823	869	861	889	876
4	790	756	871	840	853	828	846	836	859	840	885	867	890	879
5	736	699	775	754	796	767	808	772	775	744	864	846	860	846
6	504	569	647	694	643	727	658	736	604	732	754	783	740	776
7	480	550	615	666	626	635	632	631	696	648	778	765	725	755
8	401	455	469	541	452	518	447	501	429	506	634	698	652	700
9	572	515	665	627	607	591	608	585	639	588	750	734	757	738
10	628	551	753	666	819	631	816	651	762	654	830	772	828	763
11	594	591	710	725	672	738	668	728	698	723	796	793	785	797
12	502	533	638	652	550	655	560	688	566	793	758	826	735	757
13	556	523	686	674	810	658	799	656	789	590	813	773	772	759
14	499	512	545	568	522	544	518	535	529	517	653	641	688	697
15	600	507	712	586	736	680	735	640	731	633	784	739	792	752
16	544	543	638	631	612	561	595	612	646	615	761	752	746	694
17	490	451	550	522	535	556	543	562	534	543	705	715	707	703

* Estimates are multiplied by 1000.

i.e., a test on the number of significant factors. The squares of the multiple correlations of each variable with the remaining 16 variables of each set for each of the two samples were computed and used as initial estimates of the diagonal values (in the University of Illinois Illiac program). These values, recommended as lower bounds to the communalities [31], are shown in the first two columns of Table 3. Incidentally, none of the differences between the 17 pairs of corresponding multiple correlation coefficients is significant by Fisher's z transformation ($\sigma_{z_{1i}-z_{2i}} = .102$; for all pairs, $p > .05$).

The hypothesis of not more than five factors was rejected by the χ^2 test used in the Illiac program ($\chi^2 = 105.740$ for sample I and 117.698 for sample II, with critical χ^2 value of 79.9 for 61 d.f., $p < .05$). The hypothesis of only six significant factors, however, was retained since the value of χ^2 quickly dropped below the critical 5 percent value for χ^2 of 66.1 for 49 degrees of freedom (after 5 cycles, the sample I value of $\chi^2 = 56.033$ and the sample II value of $\chi^2 = 52.939$). Comparable results were obtained for Lawley's approximate test for the number of significant factors as given by Thomson [51]; these computations started from the centroid factor matrix (using adjusted high r values as estimated communalities) and used two cycles of Bargmann's procedure for determining factor loadings [4]. For both samples, the hypothesis of only five factors ($\chi^2 = 175.831$ for 61 d.f. for sample I and $\chi^2 = 160.592$ for 61 d.f. for sample II) was rejected ($p < .05$) while

the hypothesis of six factors (seventh not significant) was tenable (χ^2 = 62.319 for 49 d.f. for sample I and χ^2 = 61.394 for sample II, $p > .05$). The results of these two methods should agree since both are ways of computing solutions of Lawley's maximum likelihood equations. The second procedure illustrates, however, the usefulness and convenience of the centroid method with Bargmann's procedure for testing a given hypothesized rank.

A test of the first ten latent roots was made using Bartlett's test [5] although the results are only of incidental interest since the model being used here is *not* the principal components model. The first ten latent roots for sample I, as computed on the Illiac from the correlation matrix with unity in the diagonals, are 6.31, 2.25, 1.41, 1.27, 1.11, 0.79, 0.58, 0.49, 0.47, and 0.42. The corresponding ten latent roots for the second sample are 6.33, 2.21, 1.42, 1.14, 1.05, 0.95, 0.62, 0.51, 0.44, and 0.39. All ten roots in each sample are significant at the 5 percent level. Kaiser [37] has suggested using the number of latent roots exceeding unity as the number of factors; here that number is five, not six.

After the test of the hypothesized number of factors, the next question of the series is concerned with obtaining an objective statement of the factor pattern of rank six for the six significant factors. Three different aspects of this question can be distinguished: the estimation of communalities, the computation of the factor structure, and (for all but one pair of factor matrices) a further rotational operation. The first phase deals with estimates of the communalities to be inserted in the diagonals prior to factoring. However, in both Bargmann's and Rao's procedures, the iterated factor loadings and, therefore, the communalities, are estimated simultaneously with the test of the number of factors. Since many acrimonious and conflicting statements about the effects of differences in diagonal estimates on factor results have been made, four other sets of estimates of the communalities were computed. Because procedures for iterating communalities converge so slowly, no attempt was made to carry through the iterations to the 50 to 100 cycles that would probably be necessary to obtain convergence to four digits. However, for a given number of factors in a study satisfying the conditions for the existence of a solution, there will be a unique and determinate set of communalities [4].

A criterion of convergence was set arbitrarily at the relatively gross level of a maximum communality difference of $\pm.01$ between two successive cycles. This criterion was met for both samples after five complete cycles of the Illiac program prepared for Rao's procedure. The resulting values are shown in Table 3 in the columns headed with the abbreviations "Max. like., mult. R, (Rao)." Since Dr. Kern Dickman (personal communication) has prepared a rapid program for iterating communalities using the centroid factoring procedures, his program was used for two additional sets of estimates. The first of these started with multiple correlation coefficients and

required 15 cycles to reach the criteria of a maximum difference of .01 in communalities. As noted earlier, these multiple correlations, shown in the first two columns of Table 3, are the multiple correlations between each variable and the remaining 16 variables for each sample. The resulting iterated values are shown in Table 3 in the column headed "Centroid, mult. R, (15 cycles)." Since interest in starting with an arbitrary value such as unity in the diagonals has been expressed, Dickman's procedure was applied to this situation with the results shown in columns of Table 3 headed "Centroid, unity, (20 cycles)." These three iterative solutions seem to be approaching similar limits. Such estimated diagonals need to be compared, however, with those obtained from the widespread (and often ridiculed) practice of inserting the highest correlation coefficient or the highest residual value in each column in the corresponding diagonal cell when the centroid method is used. The results of one cycle of this successive adjustment procedure for six factors are shown in Table 3 in the column headed "Centroid, high r (adjusted)." The remaining two sets of columns in Table 3, one labeled "Prin. axes, unity" and the other labeled "Clusters, unity" are the sums of the squares of the row values in an orthogonal factor matrix of six columns and the squares of a kind of multiple correlation coefficient, both computed with unity in the diagonals of the correlation matrix by means of the principal axes and multiple-group methods, respectively.

The two analyses involving unit diagonals are *not* factor analyses. By definition the unique variances in the factor analysis model are positive and greater than zero; therefore, the diagonal values (communalities) for a factor analysis are in the range $0 < h_i^2 < 1$ [2, 4, 55]. The rotated principal axes solution with unit diagonals is a rotated principal components solution based on the first six latent vectors corresponding to the first six latent roots listed above. The cluster formulation utilizes a set of explicit objective definitions of six linearly and experimentally independent composite variables ([55], p. 63) as an illustration of one possible solution to the fourth source of confusion about factors discussed previously.

The principal axes method is well known and is unambiguous. The multiple-group cluster method, however, requires precise definitions of the clusters and linear function used. The cluster variables (composites) were defined as the average standard scores on the two or three reference tests for each factor. The variables combined are those hypothesized as associated with the PMA factor as indicated by the letter within the parentheses of Table 1. For example, the score for any individual on factor M is defined as the average of the standard scores obtained by that individual (using the means and variances of the appropriate sample) on the two variables, First Names and Word Number. For all other factors, an average of three standard scores would be used to compute (*not* estimate) the individual's factor score. These definitions are readily applied in the computations of the factor load-

ings on the normals by the multiple-group method using the sums of correlations procedure [8]. The factor loadings as computed are proportional to beta weights from the linear regression model with unit diagonals [7]. It should be noted that part-whole correlations are involved in the computations. The residuals *within* each cluster, including residual diagonals, sum to zero since these cluster vectors are group centroid vectors of the subsets of reference tests.

With the rank and the diagonal values specified, the second phase of determining the factor pattern for the six significant factors can be accomplished using the resulting covariance matrix (the correlation matrix with communalities in the diagonals). Although, theoretically, any method of factoring should be equally effective in reducing the rank of these matrices, some methods are considered as more appropriate than others as judged on the basis of efficiency or of simplicity of computations. The functions used do differ for the different methods, and these differences in method may lead to differences in the simple structure solutions. The methods used were selected, therefore, to provide data relevant to current discussions of the best method of factoring.

Eight analyses using four methods of factoring were made for each sample. These eight analyses included four complete centroid analyses, two principal axes analyses, one canonical correlation analysis, and one multiple-group analysis. The four applications of the centroid method used, as diagonal values, the one-cycle "adjusted high r" values, the 15-cycle values, the 20-cycle values, and Rao's maximum likelihood values. Both Rao's values and unity were used as diagonal entries in the principal axes analyses. Only Rao's values were used in the canonical correlation analysis. These data provide estimates of the effects upon factor loadings of three methods of factoring using a single set of diagonal values (Rao's) and of several variations in diagonals for a single method of factoring (centroid and principal axes). Only a single multiple-group analysis using diagonal values of unity was made as a demonstration of one of the many possible and simple direct solutions to a factor pattern [30]. The direct maximum likelihood solution of Howe for his Model I case ([36], pp. 82–96) was not considered for this study since invariance over diagonal estimates and method of factoring for a single rotational procedure was of primary concern.

The distributions of the 136 sixth-factor residuals from each of these applications of the four factoring methods are shown in Table 4. The means and standard deviations of the residuals are shown at the bottom of the table. Discrepancies between the distributions of residuals for the two samples are clearly shown with sample II having consistently the larger standard deviation. As one might expect, the standard deviations of the residuals computed from the "adjusted high r" centroid method are somewhat, but not markedly, larger than those obtained using the iterated communality

estimates. The distributions of residuals from the principal axes factoring method using Rao's maximum likelihood estimates have the smallest variance while the mean values are closest to zero for Rao's factoring procedure and for the centroid factoring method using the 15-cycle diagonal estimates. However, from these distributions of residuals, little preference for one method of factoring over another can be justified, even for the different sets of estimated communalities (excluding unit diagonals).

Since the approximate tests of the number of significant factors given by Burt [15], Cureton [19], and Thomson [51] are especially useful as guides in the searching, subjective, variable-defining, and hypothesis-generating process of an exploratory factor analysis, several of these tests were applied to the residuals and factor loadings of the fifth, sixth, and seventh factors of the original "adjusted high r" analysis of D. C. Bechtoldt. The data for these approximate tests are shown in Table 5. Some question as to the desirability of a seventh factor would be raised by some of these data for sample II in that original analysis. However, McNemar's test of the number of significant factors [43] based upon the ratio of the standard deviation of the distribution of residuals to the average communality agrees with the maximum

TABLE 4

Frequency Distributions, Means, and Standard Deviations of Residuals

	Centroid high r (adjusted)		Centroid . mult. R (15 cycles)		Centroid unity (20 cycles)		Centroid max. like. (Rao)		Max. like. mult. R (Rao)		Prin. axes max. like. (Rao)		Prin. axes unity		Clusters unity	
Sample:	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
Residual																
.09														1		
.08														0		
.07		2		0		0		0		0		0		0		0
.06		2		1		1		1		1		1		2		1
.05		1		1		2		1		1		0	1	1		1
.04	2	2	0	3	1	1	1	3	6	2	1	2	5	3	2	3
.03	7	4	12	3	10	5	9	2	0	6	5	7	10	9	7	5
.02	12	13	13	13	18	14	17	17	11	10	9	9	17	17	7	13
.01	24	25	25	31	16	26	24	31	25	21	32	25	17	19	26	22
00	39	29	28	38	41	38	36	36	53	59	41	45	20	24	35	36
-.01	26	22	32.	25	20	23	20	22	21	17	24	25	18	9	25	15
-.02	14	19	20	7	24	16	22	11	14	7	18	12	13	11	9	12
-.03	8	9	6	9	5	5	7	7	3	8	5	7	11	6	7	6
-.04	2	4	0	3	1	2	0	3	3	2	1	2	4	12	2	5
-.05	2	3		0		3		1		1		1	2	2	1	1
-.06		0		2		0		1		0		0	3	2	1	1
-.07		1		0		0		0		1		0	2	2	2	0
-.08													2	3	1	1
-.09													2	3	2	2
-.10 or less													9	10	9	10
Mean x 1000	-1.4	-1.7	-.02	-0.1	-0.1	-0.3	0.3	0.4	0.1	0.2	-0.7	-0.4	-14.0	-14.2	-13.8	-14.0
S.D. x 1000	17.2	22.7	16.1	18.6	16.5	18.6	16.3	18.5	15.3	17.8	14.5	16.8	44.5	45.2	42.4	45.2

TABLE 5

Data for Approximate Tests of Number of "Significant" Factors

		From previous factor		Centroid factor loadings				
						No. of loadings exceeding critical value*		
Sample	Factor being tested	Mean residual (less diag.) after sign change	Largest absolute residual	Largest (absolute)	Product two largest (absolute)	1.5σ	2σ	3σ
I	5	.028	.190	.318	.087	13	10	7
	6	.022	.109	.338	.086	10	8	2
	7	.007	.049	.186	.031	5	2	0
II	5	.041	.160	.398	.145	14	12	7
	6	.016	.126	.412	.117	8	5	4
	7	.011	.071	.268	.063	6	3	1

* Critical values based on Cureton's solution (19) of Burt's formula for σ_a. The critical values for 1.5σ are .116, .120, and .126 for factors 5, 6, and 7 respectively; for 2σ, the values are .153, .159, and .166; and for 3σ, the values are .223, .232, and .241.

likelihood procedures as to the number of significant factors; the seventh factor would not be significant by his test for any of several analyses using values other than unity in the diagonals. Since in an exploratory study, no proper test of significance of the sequential successive trial type is available, one or two additional factors might indeed be computed as suggested by Thurstone [55] and Rao [46]. Clear-cut residual planes would then aid the investigator in formulating hypotheses for a further confirmatory study using statistical tests of the hypothesized rank [4, 36].

Although the distributions of residuals were very similar from one method of factoring to another within each sample (excepting those methods using unity in the diagonals), the communality estimates shown in Table 3 did differ, especially for the two tests of factor M (rote memory factor). The discrepancies in the communality estimates for variables 1 and 2 call attention to basic and oft-repeated design requirements of factor analysis. These requirements are the necessary and sufficient conditions for identification, i.e., for a unique solution, for the case of one or more common factors as given by Anderson and Rubin [2]. Three or more variables (with nonzero elements) must be used to define a single factor in a factor analysis. (This is not a requirement, however, for the definition of factors by specified linear functions of observed variables as illustrated by the cluster solution since communality estimates are not involved.) For the case of two or more factors, three tests on each factor of a cluster configuration will satisfy the requirements. In the case of factor M, however, there are only two values which both by hypothesis and by the empirical data consistently exceed the definition used here of a zero or near zero factor loading as the range

±.10. It appears likely that this failure is responsible for the consistent large shifts in the communality estimates for variables 1 and 2 over the two samples for the three iterated sets of estimates, although smaller shifts in the corresponding estimates of other variables did occur.

Given a factor matrix from each of the several factoring operations on each of the two samples, the next phase of determining the factor pattern is to define objectively the oblique simple structure solution. Since the study was designed to provide a cluster configuration, the characteristics of the oblimax solution of Carroll [16] as modified by Pinzka and Saunders [45] should be adequate. The results of the oblimax solution expressed as factor loadings, i.e., as orthogonal projections on normals to the fitted hyperplanes, are presented for three representative factors M, V, and S in Tables 6 to 8, respectively. The three selected tables illustrate the range of variation in sampling fluctuations found in these analyses. These solutions may be termed objective ones since no change was made in any of the oblimax results of the Illiac or IBM 650 output except to define the positive direction of each normal as toward the variables with the highest factor loadings.

The oblimax solutions using only six factors can be compared on these three factors with the graphic solution shown in Table 9 and with the clusters solution given in Tables 6 to 8 in considering the fifth question of interest. The general agreement of these several solutions is clear from an inspection of the data of these tables. With an isolated configuration, the hypothesized

TABLE 6

Rote Memory (M) Factor Loadings*

Method:	Centroid high r (adjusted)		Centroid mult. R (15 cycles)		Centroid unity (20 cycles)		Centroid max. like. (Rao)		Max. like. mult. R (Rao)		Prin. axes max. like. (Rao)		Prin. axes unity		Clusters unity	
Sample: Code No.	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
1	508	463	424	757	501	640	452	717	449	713	447	714	716	708	752	730
2	638	566	820	402	715	475	766	436	763	433	765	439	838	786	808	773
3	-043	-001	-056	-048	-061	-045	-068	-051	-050	-046	-055	-049	-048	-037	-042	-005
4	054	024	025	-013	028	-011	024	-012	007	-012	022	-013	039	000	041	023
5	-019	-022	-026	-048	-020	-046	-021	-046	-004	-056	-008	-052	-009	-044	001	-018
6	-041	005	-058	-003	-057	-004	-050	002	-034	-007	-042	-001	-056	005	-044	-004
7	031	-098	031	-034	035	-021	022	-032	029	-002	029	-019	049	-057	031	-045
8	037	099	036	086	051	113	047	094	020	078	027	088	023	103	013	049
9	-103	101	-067	020	-065	033	-062	022	-074	029	-071	028	-074	056	-052	032
10	027	-078	035	-035	025	-032	026	-035	015	-030	023	-037	029	-031	009	-040
11	105	060	068	010	068	009	070	011	094	-012	085	002	079	035	043	008
12	031	044	-016	-038	-014	-034	-015	-042	005	-064	003	-050	021	012	020	005
13	033	-021	004	-001	005	014	003	013	022	019	021	011	037	026	040	017
14	-120	-049	-094	-002	-083	009	-089	002	114	035	-108	018	-119	-002	-060	-023
15	025	032	031	-011	044	-020	036	-013	034	-027	026	-017	042	-031	017	-032
16	008	079	017	168	024	200	027	171	013	182	010	175	018	162	002	103
17	017	084	004	-051	005	-064	001	-056	-008	-058	-006	-052	-002	-103	-019	-071

* Loadings multiplied by 1000.

TABLE 7

Verbal Facility (V) Factor Loadings*

Method:	Centroid high r (adjusted)		Centroid mult. R (15 cycles)		Centroid unity (20 cycles)		Centroid max. like. (Rao)		Max. like. mult. R (Rao)		Prin. axes max. like. (Rao)		Prin. axes unity		Clusters unity	
Sample:	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
Code No.																
1	021	-078	088	-080	083	-086	092	-082	072	-089	064	-092	072	-115	033	-034
2	-010	044	-041	030	-056	005	-051	019	-042	039	-033	040	-055	016	-033	035
3	624	591	580	634	562	626	566	635	600	643	593	640	675	705	694	701
4	649	601	629	625	616	630	630	629	639	633	639	633	696	685	693	680
5	590	571	610	632	625	635	608	618	586	610	589	614	696	719	687	697
6	010	-003	004	-010	006	008	025	-004	024	-005	025	-001	006	018	008	-012
7	-101	-051	-090	-007	-086	-001	-096	-005	-093	-019	-097	-019	-138	-018	-010	-038
8	140	109	152	104	163	102	167	111	151	116	154	116	218	123	092	050
9	-022	-023	-005	-036	007	-044	001	-035	-030	-018	-022	-020	-014	-025	-013	-008
10	-030	004	-024	042	-027	045	-031	042	-020	012	-025	020	-032	039	-028	009
11	035	002	046	007	056	010	051	005	046	·018	045	013	061	015	041	-001
12	038	032	022	-011	017	-003	015	-009	016	020	022	016	032	022	003	013
13	-028	-004	-035	-009	-036	-002	-038	-002	-014	-011	-021	-010	-051	-013	-047	-008
14	040	-001	046	042	051	043	053	040	034	008	040	011	068	017	045	-005
15	-028	-107	-046	-026	-045	-024	-043	-026	-043	-020	-048	-024	004	-003	-004	-037
16	066	103	069	206	075	179	064	202	052	179	051	185	105	257	068	154
17	-042	100	-044	-094	-055	-103	-052	-101	-035	-092	-036	-092	-091	-135	-064	-113

* Loadings are multiplied by 1000.

TABLE 8

Space (S) Factor Loadings*

Method:	Centroid high r (adjusted)		Centroid mult. R (15 cycles)		Centroid unity (20 cycles)		Centroid max. like. (Rao)		Max. like. mult. R (Rao)		Prin. axes max. like. (Rao)		Prin. axes unity		Clusters unity	
Sample:	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
Code No.																
1	-019	-051	-039	-018	-030	-028	-035	-024	-032	-034	-036	-029	-042	-065	-042	-052
2	081	122	063	082	063	076	060	084	061	076	064	078	082	097	043	052
3	-105	-042	-124	-032	-124	-034	-118	-030	-102	-025	-106	-028	-111	-028	-104	-035
4	-057	-032	-064	-020	-072	-020	-067	-017	-053	-018	-058	-013	-069	-024	-068	-031
5	186	063	196	081	189	087	194	082	192	075	194	080	199	089	171	066
6	006	-027	008	-039	010	-044	005	-043	002	-043	004	-040	011	-036	022	-019
7	027	036	031	061	032	061	028	057	028	062	027	056	034	070	035	077
8	-056	-031	-074	-052	-073	-055	-075	-054	-064	-048	-068	-048	-065	-057	-058	-058
9	681	708	662	676	664	664	679	674	677	679	677	679	762	785	782	796
10	822	713	873	740	872	757	849	750	835	753	838	749	885	838	885	838
11	775	765	769	775	768	770	781	767	781	765	780	766	854	823	846	828
12	038	-006	048	-038	042	-049	044	-047	051	-046	052	-044	043	-041	021	-045
13	-112	-060	-128	-057	-129	-054	-125	-047	-110	-046	-116	-047	-135	-074	-123	-064
14	096	113	118	133	118	150	122	137	126	131	126	131	133	137	102	109
15	-066	-014	-044	-016	-037	-010	-034	-011	-050	-014	-046	-011	-027	005	-018	022
16	-022	-132	013	-069	014	-067	010	-076	-013	-066	-007	-070	006	-078	010	-079
17	-009	171	016	033	017	028	019	032	004	033	008	033	-008	039	008	057

* Loadings multiplied by 1000.

simple structure for either sample is reproduced with minor variations by any of these techniques. The results of the graphic and cluster methods are consonant with those of the other methods. It should be noted, however, that the graphic solution shown in Table 9 was made in a six-factor subspace of an eight-factor (centroid) structure. Two more factors than hypothesized were computed to compensate for the inefficiency of the centroid method. The six-factor subspace was then set orthogonal to the two thinnest residual hyperplanes defined by the two principal axes corresponding to the two smallest latent roots.

The invariance of the simple structure solution over two samples of cases can be demonstrated by graphical methods, as illustrated for four representative solutions in Figures 1 and 2. Each graph contains 102 points

TABLE 9

Simple Structure Solution by Graphic Techniques

A. Factor Loadings (x 100)

Factor	M		V		W		S		N		R		Estimated Communality 8 factors	
Sample: / Code No.	I	II	I	II	I	II	I	II	I	II	I	II	I	II
1	56	55	-01	-04	-01	08	01	-08	02	-01	06	09	499	552
2	59	55	00	03	02	-03	00	09	-02	02	-03	-06	517	504
3	-07	-01	61	65	-05	-01	-09	-06	-02	-02	08	04	861	818
4	08	02	60	64	04	04	-01	-03	00	02	-06	-04	892	843
5	04	-03	44	62	03	02	10	03	01	-04	-03	01	801	760
6	-06	-06	02	-03	61	61	-01	-08	00	00	04	05	679	702
7	02	-08	-06	-06	62	61	06	08	01	-01	01	10	636	682
8	10	07	05	06	43	48	-07	-10	-01	-01	-02	-07	481	562
9	-06	-01	-09	-04	04	04	61	55	07	01	05	00	679	652
10	00	-05	01	02	05	-02	78	74	-08	00	-05	02	765	698
11	06	03	02	00	-07	02	63	75	-02	00	-01	00	733	740
12	01	-02	02	01	00	-06	00	-01	60	63	-04	-05	648	697
13	03	01	-02	-03	00	05	-07	-02	62	65	01	-05	691	689
14	-05	-02	-05	03	01	01	09	02	37	45	17	08	570	617
15	03	03	-04	05	-01	03	-05	04	07	01	51	52	716	638
16	04	22	02	25	00	-06	-02	-09	-08	00	49	35	647	658
17	00	-04	02	-05	00	17	07	05	10	-02	40	49	564	577

B. Correlations Between Primaries (x 100)*

Factor	M	V	W	S	N	R
M		44	41	07	33	41
V	40		49	06	46	63
W	48	61		10	52	49
S	09	13	13		28	26
N	37	42	38	27		60
R	35	47	42	25	56	

* Sample I values above the principal diagonal and sample II values below.

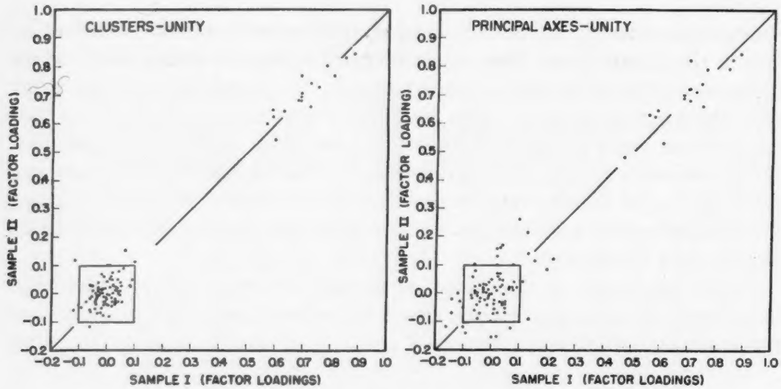

FIGURE 1
Invariance of Factor Loadings for Representative Simple Structure Solutions

(6 factors × 17 variables). The hypothesis of invariance implies that a graph of all factor loadings for one sample plotted against all factor loadings for the second sample for any one solution should show a bivariate distribution with the plotted points symmetrically placed and close to a radial line of 45 degrees. Configurational or even possibly metric invariance, as discussed by Thurstone [55], of the simple structure solution over two samples from the same population is clearly suggested by such graphical techniques for the following four solutions: the clusters with unit diagonals, the principal axes with unit diagonals, the centroid high r (adjusted), and the graphic

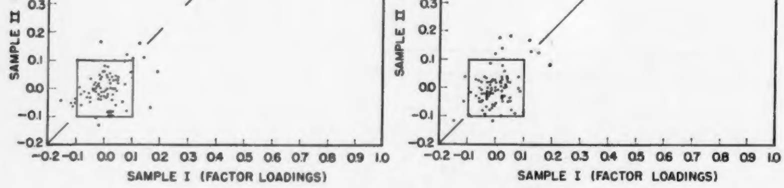

FIGURE 2
Invariance of Factor Loadings for Representative Simple Structure Solutions

(judgmental). Somewhat larger deviations for several variables from the 45-degree line will be found for the other five solutions using iterated communalities (e.g., the maximum likelihood plot in Fig. 2). The shifts for the two variables defining factor M are most conspicuous.

Figures 1 and 2 also indicate the number of nonzero loadings in each set of 102 values. Bargmann's [3] value of $\pm.10$ was used to define the zero range drawn on each graph. The clusters solution has the smallest number, 19, of nonzero values (for sample I), while the principal axes, also with unit diagonals, has the largest number, 30 (for sample II). In the graphic solution, sample I has 18 nonzero values (see Table 9). Since the hypothesized value for any one sample was 17, none of the several solutions meets this level when the data for only one sample are considered. However, two solutions, i.e., the clusters solution and the graphic solution, have only 17 *pairs* of loadings *both* greater than $\pm.10$, and the centroid high r solution has only 19 pairs with such loadings. For these three solutions, the agreement between prediction and observation is encouraging with respect both to the number of nonzero values and to the closeness of fit of the data to the 45-degree line. All of the other solutions have more than 19 pairs of loadings greater than $\pm.10$. Under certain conditions, Thurstone's concept of invariance of a simple structure solution [55] receives strong support.

Within a *single* sample, the variation in factor loadings associated with four different sets of communalities and a single factoring method (centroid) is markedly greater than is the variation associated with three methods of factoring using a single set of communalities (Rao's). These results indicate that, for a reasonably well-designed study, the centroid method is not as vastly inferior as has been suggested [37]. The effect upon the factor loadings of variation in diagonal values arising from the use of inaccurate communality estimates, however, is not the essentially irrelevant problem discussed by Wrigley [67] and Guttman [33]. They attempt to factor any arbitrary symmetric matrix and to apply to such a matrix the population rank and communality notions of factor analysis. The communality problem is a pseudo-problem unless the necessary and sufficient conditions are met for the existence of a permissible solution ([4], p. 59) to the factor analysis equations. With empirical data, the question of rank in the population is given a statistical answer under conditions for the existence of a solution.

Unfortunately, the application of available sampling formulations for the evaluation of these variations in an oblique simple structure, within a sample or between samples, either is not appropriate, or as noted by Anderson and Rubin [2], is not feasible at this time. The effect of variation in factor loadings attributable to differences in estimates of communalities (within one sample) is not a proper statistical problem since these variations represent failures to carry the iterations to convergence. Even the data from the clusters solution, however, expressed either as beta weights or, as here, as

factor loadings, cannot be evaluated in a within-sample comparison by tests of regression parameters since the factors (the independent variables) are defined by the observed test variables (the dependent variables). Such procedures as developed by Gulliksen and Wilks [28], for example, are appropriate for between-sample comparisons of beta weights or oblique projections when the separation of the independent variables and the dependent variables is maintained in the analysis. The tests of regression parameters in a single-group study are also well known for this case.

Such congruence indices as suggested by Tucker [62] and others are of little value for the matching of factors from the two samples in the current study since the congruence is so uniformly high. Instead, as descriptive statistics of congruence, the second moments (mean squares) of the differences between the pairs of corresponding columns of the rotated factor matrices as well as the second moments of the respective columns were computed. These values are shown in Table 10 for the four representative solutions exhibited in Figures 1 and 2. Tucker's index is defined as a ratio of the sum of the cross products of two columns of factor loadings to the geometric mean of the product of the sums of squares of these same two sets of values; the index, therefore, has been termed an unadjusted correlation coefficient. If desired, such congruence indices can be readily computed from the mean squares of Table 10 by means of the well-known difference formula for a correlation coefficient (without corrections for means). One of the lowest of such congruence indices, that for factor M from the maximum likelihood solution, is .826; the total congruence indices computed over 102 pairs of differences for each of these four representative solutions (in the order given in Table 10) are .987, .967, .945, and .967. Values of this order of magnitude are considered as very acceptable [62].

For any one factor, the mean squares for the sample values tend to be larger for the two solutions using unit diagonals. The larger loadings for the defining variables in these two solutions can be seen also in the figures. In addition, the mean squares of the differences are smallest for the cluster solution. These values reflect the closeness of fit of the points to the 45-degree line. For four factors, the mean squares of the differences for the centroid adjusted solution are next to the smallest although the mean squares for columns also tend to be relatively small. The largest mean squares for columns are found for factor S; this factor also has relatively small mean squares of the differences in the principal axes solution and in the maximum likelihood solution.

Comparable analyses for the several sets of data of this study indicate that differences in the stability of factor loadings do result both from the method of factoring and from the diagonal values used as communality estimates as well as from the characteristics of the data. The effects of the iteration procedures are especially evident in the mean squares of the dif-

TABLE 10

Second Moments (MS)* of Oblique Factor Loadings and of Differences Between Factor Loadings

Method of analysis	Clusters unity			Prin. axes unity			Max. like. mult. R (Rao)			Centroid high r (adjusted)		
Data / Factor	MS_I	MS_{II}	MS_D	MS_I	MS_{II}	MS_D	MS_I	MS_{II}	MS_D	MS_I	MS_{II}	MS_D
M	7.26	6.76	2.39	7.44	7.18	5.86	4.78	4.39	16.02	4.22	3.50	6.17
V	8.55	8.75	1.83	9.12	9.52	5.69	6.86	7.33	3.83	7.02	6.41	3.08
W	8.16	7.88	2.02	7.99	8.16	7.13	5.34	5.57	5.33	5.37	5.61	2.76
S	12.84	12.26	2.55	12.84	12.08	2.95	10.99	9.72	2.94	10.78	9.85	5.24
N	7.95	8.89	2.31	7.83	8.84	6.88	5.08	6.38	9.37	5.06	6.71	3.80
R	6.35	6.59	1.93	6.60	7.10	6.12	3.99	4.08	3.25	4.12	3.87	3.17

* MS_I AND MS_{II} multiplied by 100, MS_D by 1000.

TABLE 11

Congruence Indices for Arbitrary Orthogonal Factor Loadings

Method of analysis / Factors	Prin. axes unity	Max. like. mult. R (Rao)	Centroid high r (adjusted)
I	.995	.995	.996
II	.958	.942	.975
III	.662	.323	.258
IV	.824	.636	.050
V	.686	.859	.933
VI	.846	.800	.683

ferences for factors M and N. The small mean squares of the differences for the cluster solution are suggested as useful reference indices of the uncontaminated sampling fluctuations while the larger mean squares for the other three solutions include the effects of factoring method and of communality estimates. From these data as well as from similar calculations on the other solutions, one might suggest the iterated solutions fit each set of data perhaps too well from the invariance point of view, but such iterated solutions are indicated for the model under consideration.

The results of applying both the graphic and the congruence techniques to the original orthogonal factor matrices indicate little invariance for such values. The results of three such congruence analyses of representative orthogonal factor matrices are illustrated in Table 11. Tucker's index was

computed for pairs of columns of the orthogonal factor matrices ordered in terms of decreasing variance contributions of each factor. The inefficiency of the centroid method required reordering of the six "centroid high r" factors as follows: for sample I, factors 1, 2, 4, 3, 5, 6; and for sample II, factors 1, 2, 3, 5, 4, 6. No changes in order of factors were made for the other two sets of calculations, i.e., for the principal axes (unit diagonals) and for the maximum likelihood solutions. The consistency is acceptable only for the first two factors for all three analyses although some consistency is indicated for the other four factors in certain analyses. The poor showing of the centroid "adjusted high r" solution in Table 11 should be contrasted with the very acceptable degree of congruence associated with the mean squares of Table 10 and with the graphs. These data support the frequent suggestion that invariance will not be found for arbitrary orthogonal factor matrices although such invariance may be clearly indicated for a rotated simple structure solution.

The consistency indices above do not differentiate, however, between a simple structure solution and any one of the other possible factor solutions. A more direct attack on the problem of the adequacy of a simple structure solution has been made by Bargmann [3]. He considered the probability of obtaining a given number of vectors within a hyperplane section of small range ($\pm.10$) by rotational methods in a random configuration; the sampling effects (of cases) are not considered. The probability of obtaining a given frequency of zero ($\pm.10$) values of the ratios of factor loadings to length of the vectors (i.e., a_{jm}/h_j) in a random configuration was computed by Bargmann for 2 to 12 factors and for 5 to 70 variables, the range of the number of variables varying with the number of factors. For 6 factors and 17 variables (the values for the present study) Bargmann gives the number of (a/h) values in the zero range of $\pm.10$ as 10, 11, and 12 for the rejection at the 5, 1, and 0.1 percent levels respectively of the random configuration hypothesis ([3], p. 18).

The number of ratio values in the critical region for the six factors of the seven oblimax solutions and of the multiple group clusters solution are shown in Table 12. All six factors for both samples would be considered as acceptable by the simple structure 5 percent level criterion for the maximum likelihood and principal axes solutions which use the same communality estimates (i.e., Rao's) and for the multiple-group cluster solution. All but one of the other analyses had only one factor in one of the two samples with only nine ratios in the $\pm.10$ range; the "centroid high r" solution has two unacceptable planes. The number of unacceptable solutions were four for factor S in sample I, one for factor V in sample II, and one for factor R in sample II.

The relatively slight effect of factoring method on the adequacy of the simple structure can be seen in the variation in the number of zero ratios

for the three analyses using the same diagonal values (i.e., Rao's). The variations in the number of zero ratios for these three analyses represent differences of $\pm.03$ or less in the values of the ratios. The relatively greater effect upon factor loadings of changes in diagonal values are indicated by the variations among the other analyses.

TABLE 12

Number of (a/h) Ratios in Zero Range*

Method of analysis	Sample	Factor M	V	W	S	N	R
Centroid high (r)	I	12	12	13	9	13	11
	II	12	9	13	10	13	11
Centroid mult. R	I	14	11	14	9	12	12
	II	13	11	12	12	13	13
Centroid Unity	I	14	11	14	9	12	12
	II	13	10	12	12	13	11
Centroid Rao	I	14	11	14	9	12	12
	II	13	11	12	12	13	13
Max. like. Rao	I	13	11	14	10	12	13
	II	13	10	12	12	12	11
Prin. axes Rao	I	13	12	14	10	12	13
	II	13	10	12	12	12	11
Prin. axes Unity	I	14	10	14	10	12	10
	II	13	10	12	12	12	9
Clusters Unity	I	15	13	14	10	13	14
	II	14	12	13	13	14	13
Hypothesized Number		15	14	14	14	14	14

* The number of values in the "zero" range of $\pm.10$ associated with 5%, 1%, and 0.1% "probability" levels are 10, 11, and 12, respectively.

Although three solutions can be considered acceptable simple structure solutions, the data from none of these several analyses agree completely with the hypothesized number of 15 or 14 zero ratios as shown in the last line of Table 12. As noted above, the agreement is better between the number of zero factor loadings (not ratios) and the hypothesized number 15 or 14. A more adequate test of this simple structure hypothesis will probably require the use of such maximum likelihood solutions as are presented by Howe together with further developments of the sampling formulation.

The correlations between the primary axes or factors are shown for each oblimax solution and for the clusters solution in Table 13. The corresponding correlations for the graphic solution are given at the bottom of Table 9. These correlations between the factors are all positive, but the differences from sample I to sample II and from one solution to another within a sample are appreciable. The values do differ somewhat for identical diagonal values (Rao's) as a function of three methods of factoring. However, larger differences

TABLE 13

Correlations Between Pairs of Factors Defined by Primary Axes*

Methods:	Centroid high r (adjusted)		Centroid mult. R (15 cycles)		Centroid unity (20 cycles)		Centroid max. like. (Rao)		Max. like. mult. R (Rao)		Prin. axes max. like. (Rao)		Prin. axes unity		Clusters unity	
Sample:	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
Factor Pairs																
M-V	349	386	334	538	370	567	353	550	329	539	330	535	299	418	342	364
M-W	364	430	347	514	374	527	355	517	326	521	335	518	279	374	334	400
M-S	034	045	048	178	058	191	056	184	048	206	049	190	040	111	105	151
M-N	204	347	221	435	247	441	234	437	196	439	204	423	189	293	259	332
M-R	302	366	296	542	331	580	316	561	281	558	295	544	256	420	353	419
V-W	560	639	580	620	572	599	575	611	575	613	575	613	490	507	493	571
V-S	212	225	201	200	200	191	199	196	206	194	206	191	163	135	179	166
V-N	474	391	499	461	496	431	508	443	496	413	495	417	426	350	436	402
V-R	674	676	714	648	720	646	721	654	710	634	714	634	606	543	617	574
W-S	178	277	174	258	166	260	171	262	184	273	180	275	116	198	124	195
W-N	572	426	583	426	569	404	575	418	583	410	577	418	448	333	462	367
W-R	559	607	558	576	549	568	548	575	567	590	564	595	429	492	465	516
S-N	322	342	310	367	309	362	311	361	313	364	312	360	256	302	262	302
S-R	421	438	356	375	347	389	348	395	387	406	378	404	300	302	289	282
N-R	696	527	683	598	676	573	683	592	704	573	695	577	586	465	573	500

* Correlations multiplied by 1000.

are found for a single method (centroid and principal axes) as a result of changes in communality estimates; these effects on the correlations involving factor M are especially noticeable. The smallest between-sample differences were found for the cluster solutions which reflect most directly the general consistency of the original set of test intercorrelations.

It seems clear that any second- or higher-order analysis will be influenced by the diagonal values used in the first-order analysis (as well as by the number of factors and by the type of preferred solution). Invariance of second-order factor loadings can hardly be expected even from a distinctive isolated configuration unless the first-order factors are explicitly and completely defined as is the case with cluster solution. When the first-order factors are so defined, both the rank and the adequacy of the solutions of the second-order structure can be investigated as in a confirmatory first-order analysis.

The use of orthogonal simple structure [25] or of hierarchical orthogonal solutions [49, 65] does not offer any hope of greater invariance than does an oblique structure since communality estimates are involved in all of these procedures. Analytical solutions for an orthogonal structure are indeed available, but such solutions will exhibit in their first-order factor loadings a combination of the variation found here in oblique factor loadings and in correlations between factors. The forcing of orthogonality between factors in each sample (by definition) also precludes the empirical study of correlations between factors as functions of differences between treatments or

populations, differences which Anderson and Rubin [2], Rasch [47], and Thurstone [55] all noted might be associated with changes in these correlations. The regression formulation of factor analysis also indicates the irrelevance of the preference for orthogonal factors. A hypothesis of orthogonality or independence of factors in a population, of course, can be directly evaluated in terms of the correlations between explicitly defined factors in the sample.

The lack of precision of statement in the above discussion of the evaluation of sampling fluctuations is intentional. No sampling formulation for the evaluation of variations in factor loadings and in correlations between factors over both sampling fluctuations and diagonal estimates is currently available. Extensions of the work of Anderson and Rubin, Bargmann, and Howe may lead to more useful sampling formulations in the future. It is suggested that such sampling formulations for existing factor analysis models will require consideration of the several problems developed in this empirical study, i.e., the design of the study, the method for stabilizing communalities, and the method of factoring and of rotation (i.e., the specification of the properties of the preferred solution). However, the restatement of the objective of factor analysis as including the explicit definition of factors changes drastically many sampling problems. Those problems dealing with objectively defined factors are simply the usual univariate or multivariate ones. For other factor theory questions, areas of statistical theory currently under development are relevant. These areas include the identification of parameters of a structure [38] and the fitting of straight lines when both variables are subject to error [40]. When the factors are explicitly defined, these newer analytical developments also become relevant to statements about factors.

The concept of simple structure, however, warrants a brief comment. The theoretical and empirical work of Thurstone and his associates suggests the general usefulness of the concept of simple structure for the variable-defining goal of an exploratory analysis. The objective application of the concept in the current study and the results thereof indicate the possible usefulness of the concept for a confirmatory analysis. The maximum likelihood solutions using good estimators (i.e., unbiased, efficient, etc.) developed by Howe [36] make the concept a precise one. Desired analytical sampling formulations have been indicated and may eventually be developed in a usable form. For these reasons, the rejection by Maxwell [41] of the simple structure concept as not "a precise concept in a valid and efficient statistical theory of factor analysis" seems unduly severe. Under specified and attainable conditions in properly designed confirmatory factor analysis studies with zeros in designated locations, the simple structure concept of factor analysis is indeed offered as a precise concept in an incomplete but valid statistical theory.

The opinion held by Maxwell, however, can be accepted for the vast

majority of investigations entitled factor analyses and claiming to use the simple structure concept. These studies, by and large, are exploratory factor analyses (often poorly conceived) for which no statistical tests are available. Variations between investigators in the adequacy of the design of the study, in the procedures for estimating communalities, in the criteria as to when to stop factoring, and in the criteria for rotation, all create differences in the results of the factor analyses. The outcomes of these studies can be represented, at best, by lists of possible reference variables for defining an ever increasing list of factors.

But the list of possible factors is endless, or at least practically so, as emphasized by Thurstone ([54], pp. 194, 201–204, 209; [55], pp. 55–59, 62) and others, since any source of systematic differences between individuals may appear as a factor. A few of these factors, however, may indeed be selected as a stable and useful reference set of concepts accounting for most of the variance of a larger number of variables not used in the definitions of these concepts. The definition of these observable concepts by factor techniques insures some degree of linear independence among them. The usefulness of a proposed set requires in addition, however, evidence of lawful relations derived from experimental laboratory (nonfactorial) investigations of the kind recommended by Thurstone and conducted, for example, to a degree by Stukåt [50]. Starting from the available suggested definitions in, say, the ability domain [23], any investigator can provide empirical evidence as to the usefulness of these proposed definitions and of hypotheses involving them.

REFERENCES

[1] Anderson, T. W. *Introduction to multivariate statistical analysis.* New York: Wiley, 1958.

[2] Anderson, T. W. and Rubin, H. Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on mathematical statistics and probability.* Vol. 5. Berkeley: Univ. California Press, 1956. Pp. 111–150.

[3] Bargmann, R. Signifikanzuntersuchungen der einfachen struktur in der faktorenanalyse. *Sonderdruck,* Mit. d. Math. Stat., Physica-Verlag, Wurzburg, 1954.

[4] Bargmann, R. A study of independence and dependence in multivariate normal analysis. Univ. North Carolina, Inst. Statist. Mimeo. Ser. No. 186, 1957.

[5] Bartlett, M. S. Tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sec.,* 1950, **3**, 77–85.

[6] Bechtoldt, H. P. Factor analysis of the Airman Classification Battery with civilian reference tests. *HRRC res. Bull.,* 1953, 53–59.

[7] Bechtoldt, H. P. Statistical tests of hypotheses in confirmatory factor analysis. *Amer. Psychologist,* 1958, **13**, 380. (Abstract)

[8] Bechtoldt, H. P. Construct validity: a critique. *Amer. Psychologist,* 1959, **14**, 619–629.

[9] Bechtoldt, H. P. Comments on "Intraclass correlation vs. factor analytic techniques for determining groups of profiles." *Psychol.Bull.,* 1960, **57**, 157–162.

[10] Bechtoldt, H. P. and Moren, R. I. Correlational analyses of test-retest data with a thirty-year intertrial interval. *Proc. Iowa Acad. Sci.,* 1957, **64**, 514–519.

[11] Bergmann, G. The logic of psychological concepts. *Phil. Sci.,* 1951, **18**, 93–110.

[12] Bergmann, G. *Philosophy of science.* Madison: Univ. Wisconsin Press, 1957.

[13] Brodbeck, M. The philosophy of science and educational research. *Rev. educ. Res.*, 1957, **27**, 427–440.

[14] Brodbeck, M. Models, meaning, and theories. In L. Gross (Ed.), *Symposium on sociological theory*. Evanston: Row Peterson, 1958.

[15] Burt, C. L. Tests of significance in factor analysis. *Brit. J. Psychol., Statist. Sec.*, 1952, **5**, 109–133.

[16] Carroll, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

[17] Cattell, R. B. Extracting the correct number of factors in factor analysis. *Educ. psychol. Measmt*, 1958, **18**, 791–840.

[18] Cureton, E. E. The principal compulsions of factor-analysts. *Harvard educ. Rev.*, 1939, **9**, 287–295.

[19] Cureton, E. E. A note on the use of Burt's formula for estimating factor significance. *Brit. J. statist. Psychol.*, 1955, **8**, 28.

[20] Dahlstrom, W. G. Research in clinical psychology: factor analytic contributions. *J. clin. Psychol.*, 1957, **13**, 211–220.

[21] Eysenck, H. J. The logical basis of factor analysis. *Amer. Psychologist*, 1953, **8**, 105–114.

[22] Federer, W. T. Testing proportionality of covariance matrices. *Ann. math. Statist.*, 1951, **22**, 102–106.

[23] French, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monogr. No. 5*. Chicago: Univ. Chicago Press, 1951.

[24] French, J. W. Manual for kit of selected tests for reference aptitude and achievement factors. Princeton: Educ. Testing Serv., 1954.

[25] Guilford, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

[26] Guilford, J. P. Three faces of intellect. *Amer. Psychologist*, 1959, **14**, 469–479.

[27] Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

[28] Gulliksen, H. and Wilks, S. S. Regression tests for several samples. *Psychometrika*, 1950, **15**, 91–114.

[29] Guttman, L. General theory and methods for matrix factoring. *Psychometrika*, 1944, **9**, 1–16.

[30] Guttman, L. Multiple group methods for common factor analysis: their basis, computation, and interpretation. *Psychometrika*, 1952, 17, 209–222.

[31] Guttman, L. "Best possible" systematic estimates of communalities. *Psychometrika*, 1956, **21**, 273–285.

[32] Guttman, L. What lies ahead for factor analysis. *Educ. psychol. Measmt*, 1958, **18**, 497–515.

[33] Guttman, L. To what extent can communalities reduce rank? *Psychometrika*, 1958, **23**, 297–308.

[34] Henrysson, S. *Applicability of factor analysis in the behavioral sciences*. Stockholm: Almqvist and Wiksell, 1957.

[35] Holzinger, K. J. and Harman, H. *Factor analysis*. Chicago: Univ. Chicago Press, 1941.

[36] Howe, W. G. Some contributions to factor analysis. Oak Ridge National Laboratory. ORNL No. 1919, 1955.

[37] Kaiser, H. F. The application of electronic computers to factor analysis. *Educ. psychol. Measmt*, 1960, **20**, 141–151.

[38] Koopmans, T. C. and Reiersöl, O. The identification of structural characteristics. *Ann. math. Statist.*, 1950, **21**, 165–181.

[39] Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. roy. Soc. Edinburgh*, 1940, **60**, 64–82.

[40] Madansky, A. The fitting of straight lines when both variables are subject to error. *J. Amer. statist. Ass.*, 1959, **54**, 173–205.

[41] Maxwell, A. E. Statistical methods in factor analysis. *Psychol. Bull.*, 1959, **56**, 228–235.

[42] Michael, W. B. Symposium: The future of factor analysis. An over-view of the symposium. *Educ. psychol. Measmt*, 1958, **18**, 455-461.
[43] McNemar, Q. On the sampling errors of factor loadings. *Psychometrika*, 1941, **6**, 141-152.
[44] Peel, E. A. Factorial analysis as a psychological technique. *Uppsala Symp. on Psychol. Factor Analysis*, 1953, 7-22.
[45] Pinzka, C. and Saunders, D. R. Analytical rotation to simple structure II. Extension to oblique solution. Princeton: Educ. Test Serv., 1954.
[46] Rao, C. R. Estimation and tests of significance in factor analysis. *Psychometrika*, 1955, **20**, 93-111.
[47] Rasch, G. On simultaneous factor analysis in several populations. *Uppsala Symp. on Psychol. Factor Analysis*, 1953, 65-71.
[48] Reiersöl, O. On the identifiability of parameters in Thurstone's multiple-factor analysis. *Psychometrika*, 1950, **15**, 121-149.
[49] Schmid, J. and Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika*, 1957, **22**, 53-61.
[50] Stukát, K. G. *Suggestibility: a factorial and experimental analysis.* (*Acta Psychologica Gothoburgensia*, II) Stockholm: Almqvist and Wiksell, 1958.
[51] Thomson, G. *The factorial analysis of human ability.* (5th ed.) Boston: Houghton Mifflin, 1953.
[52] Thurstone, L. L. *The vectors of mind.* Chicago: Univ. Chicago Press, 1935.
[53] Thurstone, L. L. Current misuse of the factorial methods. *Psychometrika*, 1937, **2**, 73-76.
[54] Thurstone, L. L. Current issues in factor analysis. *Psychol. Bull.*, 1940, **37**, 189-236.
[55] Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univ. Chicago Press, 1947.
[56] Thurstone, L. L. and Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monogr. No. 2.* Chicago: Univ. Chicago Press, 1941.
[57] Thurstone, L. L. and Thurstone, T. G. *Primary mental abilities* (manual and tests). Chicago: Sci. Res. Assoc., 1950.
[58] Tryon, R. C. Communality of a variable: formulation by cluster analysis. *Psychometrika*, 1957, **22**, 241-259.
[59] Tryon, R. C. General dimensions of individual differences: cluster analysis vs. multiple factor analysis. *Educ. psychol. Measmt*, 1958, **18**, 477-495.
[60] Tryon, R. C. Domain sampling formulation of cluster and factor analysis. *Psychometrika*, 1959, **24**, 113-135.
[61] Tucker, L. R. The role of correlated factors in factor analysis. *Psychometrika*, 1940, **5**, 141-152.
[62] Tucker, L. R. A method for synthesis of factor analysis studies. PRS Report No. 984. Dept. Army: AGO, Personnel Res. Sec., Washington, D. C., 1951.
[63] Tucker, L. R. The objective definition of simple structure in linear factor analysis. *Psychometrika*, 1955, **20**, 209-225.
[64] Tucker, L. R. An inter-battery method of factor analysis. *Psychometrika*, 1958, **23**, 111-136.
[65] Wherry, R. J. Hierarchical factor solutions without rotation. *Psychometrika*, 1959, **24**, 45-51.
[66] Wrigley, C. Objectivity in factor analysis. *Educ. psychol. Measmt*, 1958, **18**, 463-476.
[67] Wrigley, C. The effect upon the communalities of changing the estimate of the number of factors. *Brit. J. statist. Psychol.*, 1959, **12**, 35-51.
[68] Young, G. and Householder, A. S. Factorial invariance and significance. *Psychometrika*, 1940, **5**, 47-56.

Manuscript received 6/7/60

Revised manuscript received 3/1/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

GEOMETRICAL REPRESENTATION OF TWO METHODS OF LINEAR LEAST SQUARES MULTIPLE CORRELATION

BENJAMIN FRUCHTER
THE UNIVERSITY OF TEXAS
AND
HARRY E. ANDERSON, JR.
AMERICAN INSTITUTE FOR RESEARCH

Geometrical properties and relationships of the Doolittle and square root methods of multiple correlation, as represented in the variable subspace of an orthogonal person space, are shown. The method of representation is also useful for depicting zero-order and partial correlations, as well as for the more general problem of the combination of variables.

The Doolittle [3] and the square root [7] methods of multiple correlation represent two widely used approaches to the problem of predicting values on one or more criterion variables from two or more predictor variables. The Doolittle method has been described in a number of textbooks (e.g., [20], pp. 326–331) and convenient computation forms have been devised. Dwyer [6] gives a complete proof for the Doolittle method, and Bruner [2] as well as Leavens [13], compared the accuracy of various forms of the Doolittle method.

The square root method of multiple correlation has also been presented by Dwyer [7, 8]. It is based on the square root method of factoring often attributed to Choleski [cf. 4], and it is sometimes referred to as the Gram-Schmidt method of orthogonalization. Horst [9, 10] used the essentials of the square root theory in selecting predictors for multiple criteria. Summerfield and Lubin [16, 19] used the square root method as a predictor selection device for a single criterion and made some comparisons with the Wherry-Doolittle selection method [cf. 18]. Linhart [14] described a criterion for deciding whether to use some or no predictor variables in a regression analysis. Anderson and Fruchter [1] demonstrated that the Wherry-Doolittle and Summerfield-Lubin selection methods are equivalent algebraically, computationally, and with respect to the criterion for selecting predictor variables. Both routines are based on the square root method but were shown to differ somewhat from the Doolittle method in the analytic approach to the prediction problem. The geometric representation of this difference was referred to briefly in that article. A more complete account of the geometric representation of the two approaches to multiple correlation is presented in this paper.

The three-variable case, consisting of one criterion and two predictor variables, will be used for illustration. Generalization to configurations with more than three dimensions can be made by analogy. It will be assumed that the three variables represent sets of measurements taken on a population of N individuals; the vectors representing the variables span a variable subspace within an N-dimensional, orthogonal person space. The measurements will be considered to be in deviation-from-the-mean form, a condition which places no undue restriction on the analysis. Several methods of vectorial representation of statistical concepts, developed and used by Jackson [12], Durbin and Kendall [5], Fruchter [9], Schweiker [17], and others, serve as a basis for the comparison of the multiple correlation methods.

Zero-Order Correlation

The correlation between two variables, as represented in the variable subspace of the person space, can be written as a function of the projection of the vector representing one of the variables onto the vector representing the other variable. In the flat plane spanned by two such vectors of unit length, the correlation is equal to the "line value of the rotation" of one vector from its original position to a position orthogonal to the other vector.* Unit-length vectors result where the variables are transformed to scales with unit variances.

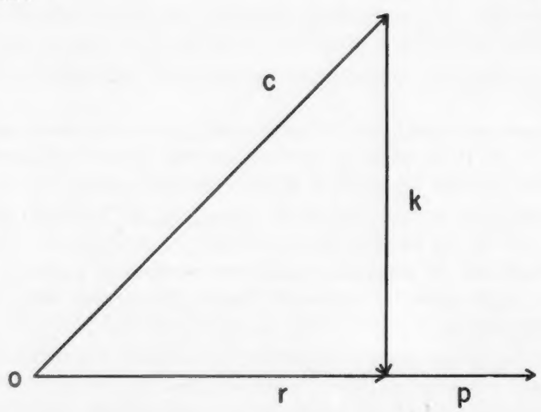

FIGURE 1

Two unit-length vectors, one representing a criterion variable c and the other a predictor variable p, are shown extending from an origin o, with c projected onto p. The vector r is the estimate component of c, while the vector k is the error component. These vectors are located in a subspace of the N-dimensional orthogonal person space, the obserservations being considered in deviation-from-the-mean form.

*The line value of rotation is defined as the projection of one vector onto another vector.

Consider a criterion vector c and a predictor vector p. The projection of c onto p, as shown in Figure 1, results in a resolution of the criterion vector into two independent components. In Figure 1, note that $c^2 = r^2 + k^2$, where c^2 is the squared length of the criterion vector, r^2 is the squared length of the estimate component of the criterion vector collinear with p, and k^2 is the squared length of the error component of the criterion vector perpendicular to p. If the N orthogonal person vectors (not shown in Fig. 1) determining the positions of the vectors representing the variables, expressed in deviation form, are divided by N, then c^2 represents the criterion's variance, r^2 represents that amount of the criterion's variance that can be estimated from the predictor, while k^2 represents that amount of the criterion's variance independent of the predictor. Two major properties of the correlation paradigm follow from the representation in Figure 1: (i) k is the shortest distance from c to p, thereby allowing r to represent the maximum amount of information in c available from p, since the errors (distributed along k) are minimized; (ii) since k is orthogonal to p, and thus to r, the error distributed along k is independent of (i.e., uncorrelated with) the estimate component.

Other properties of the zero-order correlation model in the variable subspace can be derived. By drawing a difference vector d between c and p, as shown in Figure 2, and applying a general law of cosines, the cosine of the angle θ between c and p can be written

$$\cos \theta = \frac{c^2 + p^2 - d^2}{2cp}. \tag{1}$$

Substituting in (1) the values for the lengths of the vectors and taking into

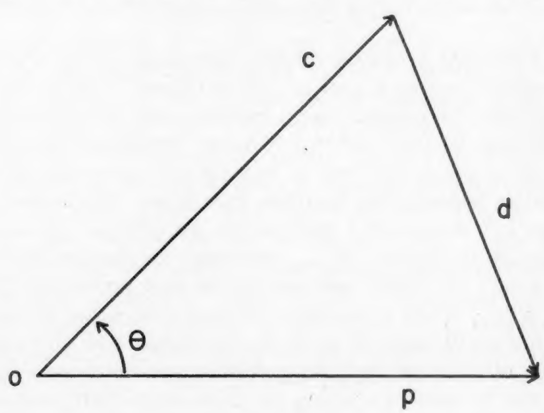

FIGURE 2

The criterion vector c and the predictor vector p located as in Fig. 1 but with a difference vector d drawn between the two vectors.

account that the orthogonal person vectors are written in deviation form (in all cases the summation is over $i = 1, 2, \cdots, N$),

$$\cos\theta = \frac{\sum x_{ic}^2 + \sum x_{ip}^2 - \sum (x_{ic} - x_{ip})^2}{2\sqrt{\sum x_{ic}^2}\sqrt{\sum x_{ip}^2}}, \tag{2}$$

which reduces to

$$\cos\theta = \frac{\sum x_{ic}x_{ip}}{\sqrt{(\sum x_{ic}^2)(\sum x_{ip}^2)}}. \tag{3}$$

Equation (3) indicates that the cosine of the angle between two vectors in the variable subspace of the person space is equal to a commonly employed definition of correlation between the two variables. Note that this result is quite general, regardless of the variance of the variables. Also, the regression weight for obtaining estimated values on c from p will be equal to the proportion of error in c accounted for by p.

From Figure 1,

$$b_{cp} = \frac{r}{p} = \frac{c}{p}\cos\theta = \frac{\sigma_c}{\sigma_p} r_{cp}, \tag{4}$$

where r represents correlation, and σ_c and σ_p the standard deviations of the criterion and the predictor variables, respectively. Where the standard deviations are equal, as in the standard score form, the regression weight is equal to the correlation.

Multiple Correlation

Figure 3 presents a variable subspace consisting of one criterion vector c and two predictor vectors p_1 and p_2. As in Figures 1 and 2, the vectors are separated so that the angular cosine between each pair of vectors is equal to the correlation between the two variables represented by those vectors.

The type of projection used in representing zero-order correlation can also be used in representing multiple correlation. The criterion vector is resolved into two independent components, an estimate component and an error component. In Figure 3, $K_{c.p_1p_2}$ represents the shortest distance between the criterion and the plane spanned by the two predictors. The estimate component $R_{c.p_1p_2}$ is not necessarily collinear with either of the predictors; but $R_{c.p_1p_2}$ lies in the plane of the predictor vectors and thus can be written as a function of these vectors. The immediate relationship to the zero-order correlation can be seen by viewing the criterion as being correlated with a dummy vector. The dummy vector, of course, is collinear with $R_{c.p_1p_2}$ and can be written also as a combination of the predictor vectors since it resides in the plane.

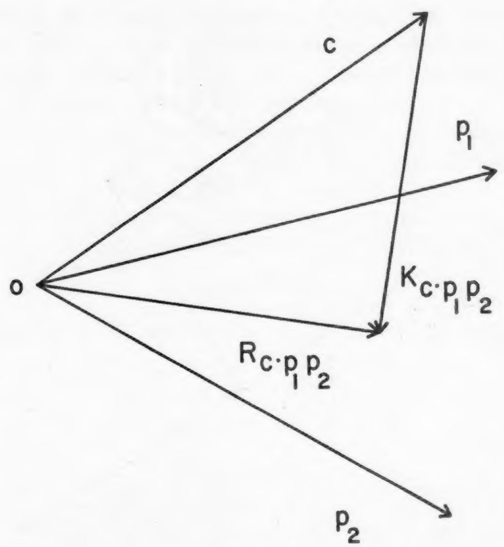

FIGURE 3

A three-dimensional space spanned by a criterion vector c and two predictor vectors p_1 and p_2 located as in previous figures, with c projected onto the plane spanned by p_1 and p_2. The vector $R_{c.p_1p_2}$ lies in the plane spanned by p_1 and p_2 and is the estimate component of c while $K_{c.p_1p_2}$ is perpendicular to the plane and is the error component of c.

Multiple Correlation Methods and Estimation of the Component Vector $R_{c.p_1p_2}$

In most presentations of the Doolittle method of multiple correlation, a set of simultaneous equations is solved to obtain the beta (β) weights for the predictor variables, which are the partial regression weights for the prediction equations where the variables are written in standard or unit variance form. The criterion and predictor variables, represented by the vectors in Figure 4, are considered to be in standard form (i.e., the observations, written in deviation form, have been divided by $\sqrt{N}\,\sigma$). It is easily shown [e.g., 17] that the beta weights for the predictors, as was true in the paradigm for the zero-order correlation, are the proportions of the predictors' unit-length vectors required to measure the estimate component vector $R_{c.p_1p_2}$. The proportions are indicated in Figure 4 by completing the parallelogram using $R_{c.p_1p_2}$ as the longest diagonal of this parallelogram. The squared multiple correlation $R^2_{c.p_1p_2}$, then, can be written

$$R^2_{c.p_1p_2} = \beta_{cp_1.p_2}r_{cp_1} + \beta_{cp_2.p_1}r_{cp_2}\,, \tag{5}$$

where r_{cp_1} and r_{cp_2} are the zero-order correlations between the criterion and the first and second predictors, respectively; $\beta_{cp_1.p_2}$ and $\beta_{cp_2.p_1}$ are the standard

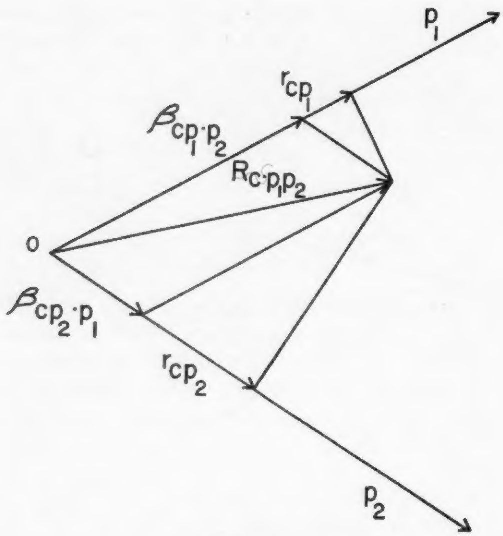

FIGURE 4

A view of the two-dimensional plane spanned by the predictor vectors p_1 and p_2 shown in Fig. 3, with the predictor-criterion correlations r_{cp_1} and r_{cp_2}, and the standard partial regression weights $\beta_{cp_1.p_2}$ and $\beta_{cp_2.p_1}$.

partial regression weights for the first and second predictors, respectively. Noting that

$$\beta_{cp_1.p_2} = \frac{r_{cp_1} - r_{cp_2}r_{p_1p_2}}{1 - r^2_{p_1p_2}},$$

and

$$\beta_{cp_2.p_1} = \frac{r_{cp_2} - r_{cp_1}r_{p_1p_2}}{1 - r^2_{p_1p_2}}, \tag{6}$$

equation (5) can be written as

$$R^2_{c.p_1p_2} = \frac{r^2_{cp_1} + r^2_{cp_2} - 2r_{cp_1}r_{cp_2}r_{p_1p_2}}{1 - r^2_{p_1p_2}} \tag{7}$$

with notation as before. Equation (7) can be used with two predictors; for more than two predictors, (5) can be expanded by adding to the right-hand member of the equation products of the predictor-criterion zero-order correlations and their appropriate standard partial regression weights.

The square root method differs from the Doolittle method in that orthogonal axes are substituted for the oblique predictor vectors. The orthogonal axes are obtained by rotating each predictor vector to a position

orthogonal to all other predictor vectors. The rotation is usually accomplished for one predictor at a time; the sequence in which the predictors are selected for rotation is arbitrary, but methods usually base the order on some relation of the predictor variables to the criterion [e.g., 10, 11, 16].

A plane determined by two predictors p_1 and p_2 , including the criterion's estimate component vector $R_{c.p_1p_2}$, is shown in Figure 5. In the figure, p_2 has been rotated to a position p_2' . The vector p_2' , as a component of p_2 orthogonal to p_1 , has the length (using the same correlational notation as in Figure 4 and assuming standard form for the variables)

(8) $$p_2' = \sqrt{1 - r_{p_1p_2}^2} \; .$$

It is obvious in this form, considering $R_{c.p_1p_2}$ as the hypotenuse of a right triangle, that

(9) $$R^2_{c.p_1p_2} = r^2_{cp_1} + (\beta'_{cp_2.p_1})^2,$$

where

$$\frac{\beta'_{cp_2.p_1}}{p_2'} = \frac{\beta_{cp_2.p_1}}{p_2},$$

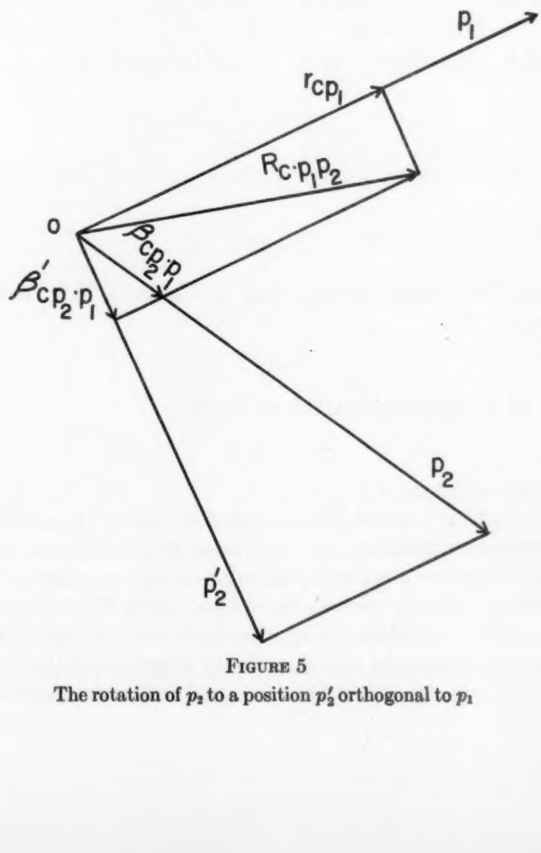

FIGURE 5
The rotation of p_2 to a position p_2' orthogonal to p_1

and $\beta'_{cp_2 \cdot p_1}$ can be defined as

$$\beta'_{cp_2 \cdot p_1} = \frac{r_{cp_2} - r_{cp_1} r_{p_1 p_2}}{\sqrt{1 - r^2_{p_1 p_2}}}. \tag{10}$$

Substituting (10) in (9) yields

$$R^2_{c \cdot p_1 p_2} = r^2_{cp_1} + \left(\frac{r_{cp_2} - r_{cp_1} r_{p_1 p_2}}{\sqrt{1 - r^2_{p_1 p_2}}} \right)^2. \tag{11}$$

Further predictors could be included in the analysis by rotating their vectors, successively, to orthogonal positions and developing additional complex terms similar to the last term in (11). Equations (7) and (11) are easily shown to be identical, and it can be seen also from the diagrams in Figures 4 and 5 that the same quantity is being estimated in both analyses.

General Properties of the Predictor Subspace and the Estimate Vector

The criterion's estimate vector can be treated as a vector representing a variable t of length $R_{c \cdot p_1 p_2}$ in the predictor-criterion configuration. From this point of view, t is perfectly predictable from the predictors and the analogue to the perfect partial correlation exists. The partial correlation between t and p_1, partialling out p_2, is usually written as

$$r_{tp_1 \cdot p_2} = \frac{r_{tp_1} - r_{tp_2} r_{p_1 p_2}}{\sqrt{(1 - r^2_{tp_2})(1 - r^2_{p_1 p_2})}}. \tag{12}$$

If p_1 and p_2 are orthogonal vectors so that $r_{p_1 p_2} = 0$, then (12) reduces to

$$r_{tp_1 \cdot p_2} = \frac{r_{tp_1}}{\sqrt{1 - r^2_{tp_2}}}. \tag{13}$$

But, similar to the rotated configuration in Figure 5, if p_1 were originally orthogonal to p_2, then

$$r_{tp_1} = \sqrt{1 - r^2_{tp_2}} \tag{14}$$

by definition of the projections involved, so that

$$r_{tp_1 \cdot p_2} = 1. \tag{15}$$

Likewise, in this situation, $r_{tp_2 \cdot p_1} = 1$.

The best possible position for the criterion vector, then, in the multiple or partial correlation problem, is in the same plane (or hyperplane) as the vectors representing the predictors regardless of the correlation between the predictors. Both problems require the measurement of the criterion's estimate vector with a combination of the predictors' vectors, although the particular vector designated as the criterion depends on the nature of the investigation and is not determined by the attendant mathematics. For

some applications, for instance, we might think of "building" a partial correlational circumstance such as the one in Figure 5, wherein p_1 and $R_{c.p_1p_2}$ are used as predictors and p_2' is designated as the criterion; here, p_1 acts as a "suppressor" since it is uncorrelated with the criterion but correlated with the other predictor. Lubin [15] has shown this as well as other properties of the partial and multiple correlational situation in algebraic terms. The geometric model points up the close relationship of the multiple and partial correlational problems.

Since the multiple and partial correlational problems involve the addition of portions of vectors to measure a component of another vector, it is easy to represent, through a geometric model, the more general problem of merely adding two variables together to form a third composite variable. This situation arises, for instance, when test subscores are added to yield a total score. The representation of this situation is shown in Figure 6 for two variables p_1 and p_2 added together to form a new variable t. As indicated in Figure 6, by completing the parallelogram and applying a general law of cosines,

$$\cos(180° - \theta) = \frac{p_1^2 + p_2^2 - t^2}{2p_1p_2}. \tag{16}$$

Substituting the N-person deviation values in (16) (all summations being over $i = 1, 2, \cdots, N$),

$$-\cos\theta = \frac{\sum x_{ip_1}^2 + \sum x_{ip_2}^2 - \sum x_{it}^2}{2\sqrt{(\sum x_{ip_1}^2)(\sum x_{ip_2}^2)}}. \tag{17}$$

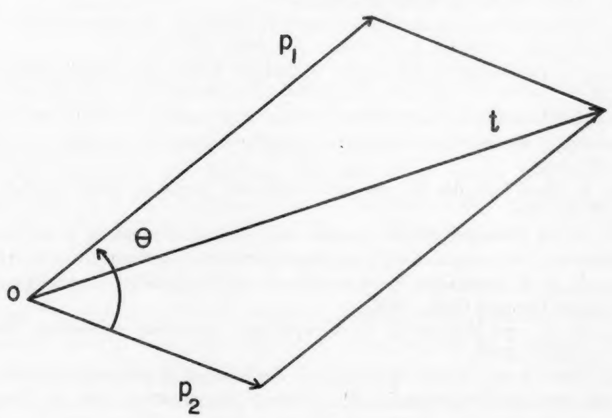

FIGURE 6

Two variable vectors p_1 and p_2 located as in previous figures, added together to form a new variable t. (The scores are in deviation-from-the-mean form.)

Noting that $\cos \theta = r_{p_1p_2}$, dividing through the numerator and denominator in (17) by N, and rearranging terms,

$$\sigma_t^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2 + 2\sigma_{p_1}\sigma_{p_2}r_{p_1p_2}\ . \tag{18}$$

The variance of the new composite variable σ_t^2 will be more a function of the variable with the larger variance as indicated in (18); also, as indicated in the flat plane in Figure 6, it will lie closer to the variable with the larger variance (i.e., the longer vector) and thus will be more highly correlated with that variable.

REFERENCES

[1] Anderson, H. E. and Fruchter, B. Some multiple correlation and predictor selection methods. *Psychometrika*, 1960, **25**, 59–76.

[2] Bruner, N. Note on the Doolittle solution. *Econometrica*, 1947, **15**, 43–44.

[3] Doolittle, M. H. Method employed in the solution of normal equations and the adjustment of a triangulation. Paper No. 3 in *Adjustment of the primary triangulation between Kent Island and Atlantic base lines*. Rep. of the Superintendent, Coast and Geodetic Survey, 1878. Pp. 115–120.

[4] Durand, D. A note on matrix inversion by the square root method. *J. Amer. statist. Ass.*, 1956, **51**, 288–292.

[5] Durbin, J. and Kendall, M. G. The geometry of estimation. *Biometrika*, 1951, **38**, 150–158.

[6] Dwyer, P. S. The Doolittle technique. *Ann. math. Statist.*, 1941, **12**, 449–458.

[7] Dwyer, P. S. The square root method and its use in correlation and regression. *J. Amer. statist. Ass.*, 1945, **40**, 493–503.

[8] Dwyer, P. S. *Linear computations*. New York: Wiley, 1951.

[9] Fruchter, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.

[10] Horst, P. A technique for the development of a differential prediction battery. *Psychol. Monogr.*, 1954, **68**, No. 9 (Whole No. 380).

[11] Horst, P. A technique for the development of a multiple absolute prediction battery. *Psychol. Monogr.*, 1955, **69**, No. 5 (Whole No. 390).

[12] Jackson, D. The elementary geometry of function space. *Amer. math. Mon.*, 1924, **31**, 461–471.

[13] Leavens, D. Accuracy in the Doolittle solution. *Econometrica*, 1947, **15**, 45–50.

[14] Linhart, H. A criterion for selecting variables in a regression analysis. *Psychometrika*, 1960, **25**, 45–58.

[15] Lubin, A. Some formulae for use with suppressor variables. *Educ. psychol. Measmt*, 1957, **17**, 286–296.

[16] Lubin, A. and Summerfield, A. A square root method of selecting a minimum set of variables in multiple regression: II. A worked example. *Psychometrika*, 1951, **16**, 425-437.

[17] Schweiker, R. F. Individual space models of certain statistics. Unpublished doctoral dissertation, Harvard Univ., 1954.

[18] Stead, W. H. and Shartle, C. L. *Occupational counseling techniques*. New York: American Book, 1940.

[19] Summerfield, A. and Lubin, A. A square root method of selecting a minimum set of variables in multiple regression: I. The method. *Psychometrika*, 1951, **16**, 271–284.

[20] Walker, H. M. and Lev, J. *Statistical inference*. New York: Holt, 1953.

Manuscript received 10/23/60

Revised manuscript received 6/9/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

A NOTE ON THE STANDARD LENGTH OF A TEST

CHARLES T. MYERS
EDUCATIONAL TESTING SERVICE

This paper describes a relationship between the variance-covariance matrix of test items and Woodbury's concept of the standard length of a test. An index of item-test relationship is described in standard length terms. The sum of these indices for the items in a test is equal to the square of Jackson's coefficient of sensitivity.

Woodbury [4] has suggested a concept which he called the *standard length* of a test—the length required for a reliability of .50. This concept has been useful in simplifying some complex problems dealing with composite scores and maximizing their reliability or validity [5]. It has been referred to in Cronbach's article [1] on coefficient alpha. Woodbury has pointed to a relationship between standard length and information.

This concept has not yet been widely used, although it seems to lead to simplicity in computations. An example of this simplicity may be demonstrated with respect to the correction for attenuation. From [4] it can be shown that the correlation between two tests of standard length equals one-half the correlation between those two tests corrected for attenuation. The validity of a standard-length test for the prediction of a perfectly reliable criterion is equal to the corrected validity multiplied by the square root of one-half. Since standard lengths are easily obtainable while infinite lengths are not, one might reasonably prefer comparisons based on standard-length tests to comparisons based on infinite-length tests. Both comparisons are equally unbiased by the various reliabilities of one's original measures.

The purpose of this paper is to show an interesting relationship between this concept of standard length and the variance-covariance matrix of test items. In order for the standard-length concept to be applied, the set of N items needed for the reliability of .50 must be representative in some sense of the test for which N is the standard length. This condition requires that the standard-length test and the other test be parallel, except for length, only to the extent that the average item variance and the average inter-item covariance of one test are the same as the corresponding averages for the other test. The specific entries may take any values allowed by this one restriction. (In test construction practice, this would often be a rigorous restriction.)

We may write the Kuder-Richardson reliability formula (20) as

$$(1) \qquad r_{tt} = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} v_i}{\sum_{i=1}^{n} C_{it}}\right],$$

where r_{tt} = the reliability of the test,
n = the number of items in the test,
v_i = the variance of item i ($i = 1, \cdots, n$),
C_{it} = the covariance of item i with the test of which it is a part. (See Gulliksen [2], p. 223, eq. 10, and note that the sum of item-test covariances equals the test variance.)

If we substitute n times the average for each of the sums, (1) becomes

$$(2) \qquad r_{tt} = \frac{n}{n-1}\left(1 - \frac{n\bar{v}}{n\bar{C}}\right).$$

Let u_i be defined as

$$(3) \qquad u_i = v_i - \bar{c}_{i.}\,,$$

where $\bar{c}_{i.}$ is the average covariance of an item with all the other items in the test, and is defined by

$$\bar{c}_{i.} = \frac{\sum_{i=1}^{n} c_{ij}}{n-1} \qquad \text{(where } i \neq j\text{)}.$$

An item-test covariance may now be rewritten as

$$(4) \qquad C_{it} = v_i + \sum_{\substack{i=1\\ i\neq j}}^{n} c_{ij} = \bar{c}_{i.}(n-1) + v_i = n\bar{c}_{i.} + u_i\,.$$

Kuder-Richardson formula (20) now has the form

$$(5) \qquad r_{tt} = \frac{n}{n-1}\left[1 - \frac{n(\bar{c} + u)}{n(n\bar{c} + \bar{u})}\right],$$

where $\bar{c}$ is the average $\bar{c}_{i.}$ for all the items in the test. Equation (5) may then be simplified to

$$(6) \qquad r_{tt} = \frac{n^2\bar{c}}{n^2\bar{c} + n\bar{u}}.$$

We may define reliability as the ratio of true variance to observed variance, and state that observed variance equals true variance plus error variance. This interpretation of reliability, if applied to Kuder-Richardson

formula (20), would be equivalent to saying: since the denominator of (6) equals observed variance, the numerator equals true variance, and the difference between the denominator and the numerator, $n\bar{u}$, equals error variance. The application of the Spearman-Brown prophecy formula to Woodbury's concept of standard length requires that $\bar{u}$ be a constant for any length test to which the formulas apply.

Solving (6) for N, the number of items when $r_{tt} = .50$,

$$N = \frac{\bar{u}}{\bar{c}}. \tag{7}$$

The ratio of average error (or noise) per item to the average inter-item covariance is equal to the number of items required for a standard-length test and is a constant for any specified set of items, regardless of test length within the restrictions applicable to the concepts we have been discussing.

Using these terms, one can write an index of item-test relationship that describes the "length" of a single item in standard-length units. This index would be defined by

$$k_i = \frac{\bar{c}_i}{\bar{u}}. \tag{8}$$

Note that any index of item-test relationship, whether it be an item-test correlation or an estimate of true and error variance in an item, describes an item in context; item indices are never completely independent of the context. Such statistics are dependent upon the population used in collecting the data and the set of items used as a criterion.

When the values of k_i are summed over all the items in a test, the sum equals the length of the test in standard-length units. Therefore, if K is defined as the sum of the k_i values for all the items in a test, the reliability of the test, using the Spearman-Brown formula, is

$$r_{tt} = \frac{K(.50)}{1 + (K - 1)(.50)}. \tag{9}$$

This reduces to Woodbury's [4] equation

$$r_{tt} = \frac{K}{K + 1} \quad \text{or} \quad r_{tt} = \frac{\sum k_i}{1 + \sum k_i}. \tag{10}$$

This may be an interesting way of describing the relationship between items and the reliability of a test.

The quantity K is equivalent to the square of Jackson's [3] coefficient of sensitivity (γ), just as Kuder-Richardson reliability is equivalent to ρ in Jackson's article. The standard-length concept presumably shares the advantages that "sensitivity" has for describing a test.

REFERENCES

[1] Cronbach, L. J. Coefficient *alpha* and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.
[2] Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.
[3] Jackson, R. W. B. Reliability of mental tests. *Brit. J. Psychol*, 1939, **29**, 267–287.
[4] Woodbury, M. On the standard length of a test. *Psychometrika*, 1951, **16**, 103–106.
[5] Woodbury, M. and Lord, F. M. The most reliable composite with a specified true score. *Brit. J. statist. Psychol.*, 1956, **9**, 21–28.

Manuscript received 3/28/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

A COSINE APPROXIMATION TO THE NORMAL DISTRIBUTION

DAVID H. RAAB AND EDWARD H. GREEN
BROOKLYN COLLEGE

A cosine function is suggested to approximate the normal distribution as a device for simplifying algebraic manipulations of the latter. Numerical evaluations remain straightforward and employ only the commonly available trigonometric tables. A method of visual curve fitting requiring only an oscilloscope is also described.

The purpose of this note is to call attention to a trigonometric approximation to the normal distribution which may be of value in psychological investigations. The substitution was suggested by one of us (EHG) for use in a statistical model being developed by the other (DHR).

The function,

$$f(x) = \frac{1}{2\pi}(1 + \cos x), \qquad (-\pi < x < \pi) \tag{1}$$

has a symmetric, bell-shaped graph with mean at $x = 0$ and with unit area. Its variance is given by the expression,

$$\sigma_x^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2(1 + \cos x)\,dx, \tag{2}$$

which is easily evaluated. Its standard deviation is $\sqrt{(\pi^2/3) - 2}$ radians, which is approximately 1.14 radians. Since the distribution extends from $-\pi$ to π, it includes approximately $2\pi(.88) = 5.53$ standard deviations.

In Fig. 1, the normal probability density function and one cycle of the cosine function are plotted on the same coordinates. As drawn, they both have unit sigma and unit area. The cosine is obviously platykurtic as compared to the normal distribution, but the fit is not outrageously poor.

The trigonometric function has several features which commend it. Analytical expressions which involve products or powers of the normal density function and its integral are not easily evaluated. Replacing the normal function in such expressions by the cosine often yields expressions which can be evaluated algebraically. To be able to manipulate theoretical equations at one's desk is of great value.

As an example, consider two identical normal density functions, $\phi(x)$. If these are distributions of two independent variables, the joint sampling distribution $F(x)$ of whichever variate is smaller is

$$F(x) = 2\phi(x)\int_{x}^{\infty} \phi(x)\,dx. \tag{3}$$

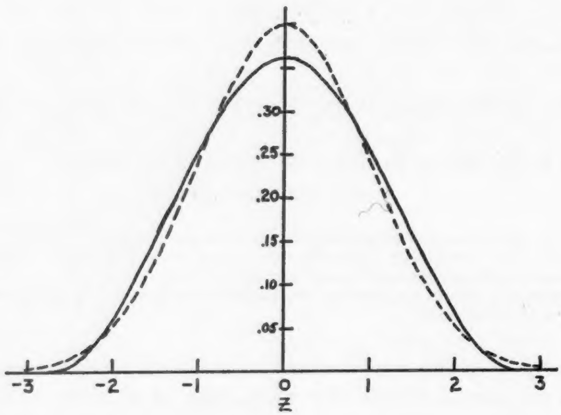

FIGURE 1

The normal distribution (dashed line) and the cosine function (solid line) drawn to have the same mean, unit area, and unit sigma.

Since $\int_{-\infty}^{\infty} F(x)\,dx = 1$, the mean of $F(x)$ is given by the expression,

$$(4) \qquad \overline{F(x)} = \int_{-\infty}^{\infty} xF(x)\,dx = 2\int_{-\infty}^{\infty} x\phi(x)\left[\int_{x}^{\infty} \phi(x)\,dx\right] dx.$$

This mean is readily found if $\phi(x)$ is replaced by

$$f(x) = \frac{1}{2\pi}(1 + \cos x),$$

and the limits of integration are adjusted to take account of the fact that $f(x)$, from (1), exists only between $-\pi$ and π. Equation (3) becomes

$$(5) \qquad F(x) = 2f(x)\int_{x}^{\pi} f(x)\,dx,$$

which reduces to

$$\frac{1 + \cos x}{\pi} \cdot \frac{\pi - x - \sin x}{2\pi}.$$

From (4), $\overline{F(x)}$ is then readily evaluated to be $(-\pi/3 + 5/4\pi)$ which equals -0.649 radians or -0.572 sigmas. Equation (4) was subsequently evaluated by numerical methods (Gauss quadrature) with the help of an IBM 7090. The result obtained for $\overline{F(x)}$ was -0.564 sigmas.

Numerical evaluations employing the cosine distribution remain simple and straightforward. Ordinates and areas of the function are easily derived from tables of the cosine and the sine. Raw scores are first transformed into

standard form and then into radian or degree units. (Recall that $\sigma = 1.14$ radians and that one radian $= 180/\pi$ degrees.) As an example, Table 1 was prepared from trigonometric tables by evaluating expression (1) to give the ordinates in column 3 and using

$$\frac{1}{2\pi}\int_0^x (1 + \cos x)\,dx = \frac{1}{2\pi}(x + \sin x)\Big|_0^x \tag{6}$$

to give the areas in column 4.

TABLE 1

Ordinates and Areas of the Unit Cosine Distribution

z	x (radians)	f(x)	Area from mean to x
0.0	0.00	.318	.000
.2	.23	.314	.073
.4	.46	.302	.144
.6	.68	.283	.208
.8	.91	.257	.271
1.0	1.14	.226	.326
1.2	1.37	.191	.374
1.4	1.60	.156	.414
1.6	1.82	.119	.444
1.8	2.05	.086	.468
2.0	2.28	.056	.484
2.2	2.51	.032	.495
2.4	2.74	.012	.496
2.6	2.96	.003	.499
2.76	π	.000	.500

An interesting by-product of the trigonometric substitution is that approximate curve fitting can be carried out employing only an oscilloscope. Whereas Gaussian plots are expensive to generate, the AC power line provides a sinusoid of sufficient purity for our purpose.

The procedure consists of displaying one cycle (from trough to trough) of the line signal and fitting this to a frequency graph of the data in question.

The histogram (drawn on fairly thin graph paper) is held against the face of the oscilloscope, and the vertical gain, horizontal position, and sweep time (or horizontal gain) controls are adjusted for best fit. The three parameters of the distribution are at once available. The mean is the abscissa value under the peak of the sinusoid, N equals 0.88π times the peak-to-trough amplitude, and sigma is easily calculated from the wavelength of the display. (Recall that one wavelength equals 5.53 sigma.)

Conversely, the best fitting cosine can be generated from the sample statistics by reversing the three relations. To be specific, the cosine wavelength is set equal to the given standard deviation multiplied by 5.53, and the pattern is centered so that its peak occurs at the given mean position. Finally, the peak-to-trough height is adjusted to equal N divided by 0.88π (i.e., 0.36). The resulting pattern has the relation to the corresponding Gaussian distribution that is shown in Fig. 1.

Manuscript received 10/10/60

Revised manuscript received 3/6/61

PSYCHOMETRIKA—VOL. 26, NO. 4
DECEMBER, 1961

A RETRACTION ON INTER-BATTERY FACTOR ANALYSIS

W. A. GIBSON
DEPARTMENT OF THE ARMY*

In [2] there was a critique of Tucker's inter-battery factor analysis [4] whose main point was that, in general, Tucker's procedure does not provide orthogonal factor matrices, i.e., matrices whose elements are interpretable as correlations between tests and uncorrelated factors. That point remains correct. However, a valid post-publication criticism of [2] has been made in a letter to the author from Carl Bereiter of Vassar College. To illustrate the main point of [2], there was included a fictitious example so constructed that the orthogonal factors necessary to account for the inter-battery correlations R_{12} also accounted completely for the within-battery correlations R_{11} and R_{22} . Application of Tucker's solution exactly as he outlined it ([4], pp. 118–119) would have led to discrepancies in reproducing the within-battery correlations ranging in absolute size from .08 to .10, thus verifying that the inter-battery method had not yielded orthogonal factor matrices. However, Tucker's recommended procedure was not followed for reasons of symmetry ([2], equations (1) and (2), p. 20, and Table 2, p. 21); this had a dramatic result which Bereiter has shown can always be avoided. Imaginary factor loadings emerged, along with excessively poor reproduction of the within-battery correlations. In the symmetric approach taken in [2], two roots-and-vectors resolutions were employed, rather than the single one proposed by Tucker, and it was not then recognized that merely coordinating the essentially arbitrary directionalities of the two sets of characteristic vectors would have eliminated the imaginary loadings. Thus a retraction is required on published statements [1, 2] to the effect that imaginary loadings are sometimes unavoidable in inter-battery factoring.

There remains the possibility, perhaps more theoretical than practical, that a partially imaginary inter-battery solution could better reproduce one or both sets of within-battery correlations than could an entirely real solution. That would create a dilemma analogous to the Heywood case (cf. [3], pp. 289–290), especially when, as will often be true, inter-battery factoring is used only as a means to the end of avoiding communality estimation while trying to fit all side correlations, rather than as an end in itself—that of exclusive concern for between-battery factors. Here the dilemma is whether to fit best or to fit with real loadings only. Such a dilemma

*Opinions expressed herein are the author's, not the Army's.

is not unique to inter-battery factoring, however, for in regular factor analysis cases can be constructed where the side correlations are "explained" with lower rank if imaginary loadings are allowed.

A reviewer-proposed modification of Tucker's procedure to yield proper orthogonal factor matrices is discussed at the end of [2]. Another solution is presented in [1] and [1a].

REFERENCES

[1] Gibson, W. A. An expansion upon Tucker's inter-battery factor analysis. *Amer. Psychologist*, 1960, **15**, 487. (Abstract)

[1a] Gibson, W. A. An asymmetric approach to multiple-factor analysis. *Brit. J. statist. Psychol.*, 1961, **14**, Part II.

[2] Gibson, W. A. Remarks on Tucker's inter-battery method of factor analysis. *Psychometrika*, 1960, **25**, 19–25.

[3] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.

[4] Tucker, L. R. An inter-battery method of factor analysis. *Psychometrika*, 1958, **23**, 111–136.

Manuscript received 3/1/61

Revised manuscript received 6/17/61

BOOK REVIEWS

HARRY H. HARMAN. *Modern Factor Analysis*. Chicago: University of Chicago Press, 1960. Pp. xvi + 469.

During the last twenty years remarkable advances have been made in what is commonly called factor analysis. In 1940 it was still regarded both by many psychologists and statisticians as a rather rough and ready substitute for systematic observation and experiment, with no very sound mathematical basis and productive of somewhat questionable psychological doctrines. Today it is widely used, not only by psychologists in many different fields but also by research workers in other branches of science. The change, as Harman points out, is due partly to the advent of the electronic computer and partly to the advances made in the general theory of multivariate analysis.

Harman's book is primarily a revision of the textbook jointly compiled by Holzinger and Harman in 1941, and then entitled *Factor Analysis: A Synthesis of Factor Methods*. But nearly every chapter has been rewritten; much has been added, and much left out. Like that earlier work, it endeavors to set forth in lucid and unambiguous terms first the logical foundation of the methods described, and then the mathematical formulas on which they are based. The theoretical developments are supplemented by detailed computing procedures, and illustrated by numerous well-chosen examples. A welcome feature is the attention paid to a clearly defined terminology and notation, which will unquestionably help to provide a standardized set of concepts and symbols.

The general lay-out remains much as before. The main text is subdivided into four parts. The first deals with the "foundations of factor analysis," and the changes introduced make for much greater clarification. It starts with a brief history of the subject, which it dates from Karl Pearson's "crucial article" of 1901 on the method of "principal axes." This is followed by a introductory chapter on the factor analysis model. This is assumed to be linear, and is described as including (i) general factors (positive or bipolar), and/or (ii) group factors, and (iii) unique factors. A careful distinction is drawn between the factor pattern (which furnishes a regression equation of the classical type) and the factor structure, which, it is contended, should also be given explicitly (except of course when the two are identical, as with uncorrelated factors). The chapter on the relevant geometric concepts remains much as before; but the discussion of the relevant matrix concepts is now removed from the appendix to the text. The square-root method for solving systems of linear equations (illustrated by a worked example) is recommended for desk calculations as being easier and more compact than the more familiar Doolittle. The discussion of the determination of communalities is considerably expanded to include a wider variety of theoretical, approximate, and arbitrary procedures; and the chapter on preferred types of orthogonal solution is transformed into an exposition of the properties of different types of solution.

Part II, as before, deals with direct solutions. To the previous discussion of the two-factor, bifactor, centroid, and principal factor solutions there is now added a new chapter on the multiple-group solution. Part III, on derived solutions, has been almost wholly rewritten. In Harman's view one of the most important advances during the past twenty years has been the development of "objective procedures for determining simple-structure solutions." Two new chapters have therefore been inserted to incorporate modern analytical methods of rotation—the quartimax and varimax for the orthogonal case, and the oblimax, quartimin, and oblimin for the oblique case. This comprehensive survey of the newer techniques and the newer computing procedures including recent work by Carroll, Dwyer, Kaiser, Neuhaus, Saunders, and Wrigley, hitherto accessible only in scattered periodicals or research reports,

forms one of the most useful features of the book. Part IV is chiefly concerned with two special topics, first the measurement of factors and second the available statistical tests for factorial hypotheses. For the latter—a valuable addition to the volume—Ardie Lubin prepared the first draft. Part V is entirely new. It consists of excellently chosen problems and exercises on each of the chapters with fully worked answers; but the rest of what is still essential in the original appendices has now been incorporated in the text. The book ends with an extensive bibliography of over 400 entries.

Many of the foregoing changes, together with the discussions of preferred methods, appear to reflect a change in outlook. In the earlier volume nearly fifty pages were devoted to methods for obtaining the type of factor matrix which includes both a general and a set of nonoverlapping group factors, typified by Holzinger's own bifactor method. In the revision the chapter on the bifactor method has been reduced to 27 pages, while the discussion of methods for obtaining simple structure is expanded to over a hundred. The procedures which now seem to be preferred are no doubt more rigorous and more amenable to exact statistical treatment. But the general effect, I fancy, may be to exclude from the reader's consideration a wide variety of factorial hypotheses, which have proved of frequent occurrence, not only in psychology, but also in other fields of research. If, to borrow Harman's term, we regard factors as specifying a set of categories or principles of classification, then one of the commonest schemes will naturally be that which embraces first a generic factor, and then a series of group factors, where the broader groups, like the initial general factor, are subdivided into narrower groups. An urgent need therefore is a method of analysis which like Thurstone's includes overlapping group factors, and like Holzinger's a general factor, and at the same time shall have the mathematical rigor of (for example) Lawley's method of maximum likelihood. However, until some such procedure has been worked out, most research workers will probably prefer those which Harman has so fully and fairly discussed. As he says, "the heated controversies about the best procedure are over; each has had its place in the growth of the method."

His book sets out to deal with methods only. Certain topics which sometimes find places in books on factor analysis have therefore been rejected. Nothing is said about the practical application of factor analysis either in psychology or in other fields of work; and no attempt is made to summarize the concrete results which psychologists have so far obtained. A few special modifications are referred to at the very close, but not explained or discussed, for example, the so-called inverted techniques, where the correlations to be factorized are correlations between persons rather than between tests, traits, or other attributes. Nor are any examples given of factorial methods as applied to the analysis of qualitative data. Guttman's earlier work on methods of assessing communalities and of factorization by multiple group are duly mentioned; but his modifications of factorial techniques are passed over in silence.

But it is always unfair to criticize, or seem to criticize, a volume for not taking up problems which the author, for excellent reasons of his own, deliberately excluded from his program. There was a manifest need for just such a textbook as this; Harman has admirably fulfilled it. The volume will at once become one of the standard works on factorial methods, and will prove indispensable to every student who contemplates using factor analysis in his own researches.

University College, London CYRIL BURT

EMANUEL PARZEN, *Modern Probability Theory and Its Applications*. New York: John Wiley and Sons, Inc., 1960. Pp. xv + 464.

The need for a basic course in probability theory as part of the background for advanced courses in such diverse fields as "statistics, statistical physics, industrial engineering,

communication engineering, genetics, statistical psychology, and econometrics" is becoming more and more widely recognized. Accordingly, the author has set himself the task of writing a textbook "that can be adapted to the needs of students with diverse interests and backgrounds." I shall attempt, in the following, to assess the degree to which he has succeeded in this task.

The first six chapters can, the author believes, be handled by students who have had only one year of college calculus. The titles of these chapters are: "Probability Theory as the Study of Mathematical Models of Random Phenomena," "Basic Probability Theory," "Independence and Dependence," "Numerical-Valued Random Phenomena," "Mean and Variance of a Probability Law," "Normal, Poisson, and Related Probability Laws." These are sufficient, it is claimed, to "constitute a one-quarter course in elementary probability theory at the sophomore or junior level." The next two chapters, "Random Variables" and "Expectation of a Random Variable," require somewhat more mathematical background, while the last two, "Sums of Independent Random Variables" and "Sequences of Random Variables," are "much less elementary in character than the first eight chapters."

To show the wide applicability of probability theory, the author presents a large array of examples and exercises relevant to each of the fields cited above (and more). He also emphasizes, throughout the text, the notion that probability theory is a *mathematical model* of real phenomena, or, more specifically, an axiomatic theory dealing with "those methods of analysis that are common to the study of random phenomena in all the fields in which they arise." He thus repeatedly points out what probability theory can, and what it cannot, do. It cannot, for instance, "give rules for the construction of sample description spaces." Such rules belong in "the art of applying the mathematical theory...to the study of the real world." The author's remarks in connection with Bayes' theorem and Laplace's rule of succession (pp. 119–124) are especially pertinent. One sometimes hears assertions to the effect that, because their applications often lead to absurd results, the validity of these theorems is questionable. Such assertions reflect a misconstruing of the nature of mathematical theories. As Parzen correctly points out, the theorems are strictly valid within the axiomatic system. What is at fault, when absurd results are obtained, are the "coordinating definitions" that identify some of the terms of the mathematical theory with aspects of the observable phenomena.

In one instance, however, Parzen seems to lose sight of the dictum of separating model from phenomenon. With reference to Bernoulli's law of large numbers (p. 229), he states that "one can reach a conclusion *within* the mathematical theory of probability that may be interpreted to *justify* the frequency interpretation of probability..."(italics mine). Surely this cannot be the case. Just as certain absurd results do not invalidate Bayes' and Laplace's theorems, so likewise is it impossible for the analytic law of large numbers logically to validate the empirical frequency concept of probability. In fairness to the author, it must be mentioned that he does describe the relationship between the analytic and the empirical laws of large numbers more correctly in Chapter 10 (p. 417). But not all students are presumed to get up to this point in the book.

Next, with regard to the author's aim of making his book suitable for students with varying degrees of mathematical background, one cannot object to his graduating the chapters in terms of mathematical sophistication in the manner earlier cited. One may, however, question whether he has incorporated enough topics within the first six chapters ("one-year calculus" level) for a self-contained introductory course in probability theory, and also whether the treatment in these chapters is always appropriately geared to the intended level. My impression is that the answers to both these questions range from "dubious" to "no."

The postponement of all but a cursory mention (on p. 238) of the central limit theorem until Chapter 8 is an example pointing to the insufficiency of the minimal block comprising Chapters 1 through 6. Also, the absence, within this block, of any discussion of how to

derive the distribution of the sum of two random variables would seem to make it less than complete even for an introductory course.

As to the second question, there are marked fluctuations in the amount of mathematical sophistication which the author presumes a student would possess after one year of calculus. For instance, the notion of difference equations is casually introduced as an exercise in Chapter 3, Section 4, and is then heavily used in the next two sections dealing with Markov chains. Again, the technique of differentiation under the integral sign is used, without prior explanation, in connection with moment generating functions (p. 216). On the other hand, seven pages (pp. 35–41) are devoted to the exposition of elementary combinatorial algebra and the binomial and multinomial theorems—topics with which, theoretically, any student who has had college algebra should be familiar. If the author deems it necessary to review these topics at such length, one would expect him to feel even more obliged to give fairly detailed discussions of some of the less elementary techniques such as those mentioned above.

In the remainder of this review, I list a number of minor defects ranging from factual errors to points of poor style or lack of clarity. Typographical errors (which are not excessive in number, considering that this is a first printing) are not included here.

P. 62: The condition preceding equation (4.5) should be that $1 > P[A] > 0$ (rather than just $P[A] > 0$), since the equation involves $P[B \mid A^c]$, which, according to (4.4), is undefined if $P[A^c] = 0$—as it would be if $P[A] = 1$.

P. 81: The statement, "This is so if and only if, for some $j = 1, \ldots, n$, s does not belong to A_j, which is equivalent to, for some $j = 1, \ldots, n$, $I(A_j;s) = 0$, which is equivalent to..." is in very poor style: a confusion between two levels of language.

P. 140; first line: The statement, "If all states in a Markov chain communicate...," should be corrected to read, "If all *pairs of* states ...."

P. 184: Equation (5.1) fails to specify the value of $P[B]$ in the case when B is not a subset of S but yet has a nonempty intersection with S.

Pp. 201–202: It is misleading to speak of the absolute dispersion, the square dispersion, and the third central moment as being among the "many kinds of averages one can define" for a set of data. To be sure, they are averages of different functions evaluated for numbers in the data set, but they are so by virtue of a prior definition of a unique average for any set of numbers—otherwise, we would have an infinite regress of sorts.

P. 214: The nomenclature "p percentile" (where p is a number between 0 and 1) is incongruous; it should be "100 pth percentile (or *centile*)."

P. 226: The statement, "This lower bound, known as Chebyshev's inequality, was named after ..." obviously involves an incorrect apposition. It should be amended to read, "This lower bound is given by Chebyshev's inequality, a proposition named after ..." or something like that.

P. 303: The upper part of Fig. ~~70~~ 7c is incorrectly drawn. The dotted line whose length is Y_2 should be perpendicular to the chord, and Y_1 should be the angle between the extension of the chord and that of the radius shown in dotted line.

P. 346: The author's objection to the terminology, "*expected value* of the random variable X," as being "somewhat misleading" is itself misleading. He would have $E[X]$ designated as "the expected value of the arithmetic mean of a random sample of the random variable." But a single value of a random variable *is* the arithmetic mean of a random sample (of size one) of the random variable!

P. 363: The statement, "The correlation coefficient provides a measure of how good a prediction of the value of one of the random variables can be formed on the basis of an observed value of the other" obviously needs to be qualified by mentioning *linearity* somewhere. (This point is later clarified, on p. 387.)

Finally, there are two sources of minor annoyance that run throughout the book. One

is the inordinate frequency with which example problems anticipate material discussed later. (Without attempting to be exhaustive, I counted 13 such instances.) The other is that references are cited, with complete publication data, in the running text. This practice not only disturbs the continuity of the text, but makes it difficult to locate the references. A bibliography at the end of each chapter, or at the end of the book, would be helpful.

Despite the several shortcomings noted in the foregoing, I believe this book is a valuable addition to the textbook literature in the field of probability theory. Its presentation of a rigorous axiomatic approach, at a level by and large suitable for the undergraduate, is admirable. Thus, provided the instructor is willing to supplement the mathematical information at certain points, he should find this a highly usable text, suitable for students who are preparing themselves for any of a large variety of fields, including psychometrics and other areas of quantitative psychology.

University of Hawaii at Hilo MAURICE M. TATSUOKA

H. GULLIKSEN AND S. MESSICK (Eds.). *Psychological Scaling: Theory and Applications.* New York: John Wiley and Sons, Inc., 1960. Pp. xvi + 211.

This symposium derives from a conference on psychological scaling held at Princeton, New Jersey, in May 1958. The topics with which it deals extend much beyond the range usually included under the title of "psychological scaling." From this point of view the title of the monograph is too limited. The prevailing contribution is theoretical with only suggestions of how developments can be applied. From this point of view the title is too broad.

The participant authors will generally be recognized as outstanding contributors in connection with their special problems by all readers who have followed publications in scaling and related subjects. One session dealt with "scaling properties," with papers mostly pertaining to interval scaling, authored by Lyle V. Jones, Warren S. Torgerson, Roger Shepard, and Bert F. Green, Jr. Shepard considers some new theory basic to scaling derived from discrimination behavior.

In the section on psychological scaling, S. S. Stevens marshalls much evidence in support of the power psychophysical law and for the preference of ratio scaling over discrimination and interval scaling. The same information may be found in Stevens' other numerous publications, but it is succinctly summarized in his chapter. William McGill presents ratio-scaling data from judgments of loudness, data that pose some theoretical and methodological questions.

In a session on test theory, Paul F. Lazarsfeld offered some new developments linking latent structure analysis with problems of test-item theory. Frederic M. Lord presented further thinking on true scores for individuals and how they may be inferred from obtained scores.

The subjects of utility theory and measurement and decision making receive treatment at the hands of Ward Edwards, Sidney Siegel, and R. Duncan Luce. The relations of these subjects to scaling are becoming clearer.

Multidimensional scaling models are treated by Clyde H. Coombs, Ledyard R Tucker, and Robert P. Abelson. Tucker makes explicit comparison between intra-individual and inter-individual multidimensionality.

A single bibliography includes more than 150 titles. Unlike some monographic volumes, this one contains a useful index.

This volume would be a good place for a nonpsychologist with advanced mathematical training to find out how quantitative psychology is developing in most of its aspects. The book is not for the average psychologist who does not share such a mathematical back-

ground or who has not previously followed the theoretical developments included. It does bring together the recent (up to 1958) thinking of a number of leaders in quantitative theory and methods and represents the frontiers of thinking in those subjects.

University of Southern California J. P. GUILFORD

Minutes of the

1961 ANNUAL BUSINESS MEETING

of the

PSYCHOMETRIC SOCIETY

The regular Annual Meeting of the Psychometric Society was held in New York, N. Y., on Wednesday, 6 September 1961. President John B. Carroll called the meeting to order at 10:00 a.m.

The minutes of the previous Annual Meeting were approved.

Dr. Robert L. Ebel reported for the Membership Committee. The Membership Committee nominated 78 persons as full members, and 22 individuals as student members.

It was moved, seconded, and passed that the following 78 persons be elected as full members.

Gonzalo Adis-Castro
Harry Edwin Anderson, Jr.
Norman H. Anderson
Alexander W. Astin
Daniel E. Bailey
Joan Hauser Bailey
Thomas J. Banta
Ernest Stoelting Barratt
Albert E. Beaton
Robert Besco
Harold F. Bligh
Clarence Bradford
N. Philip Bryden
Robert R. Bush
T. Anne Cleary
Raymond O. Collier, Jr.
William W. Cooley
Norman A. Crowder
Fred L. Damarin
Herbert A. David
Joseph R. Devane
Jean Eugler Draper
Doris Entwisle
E. V. Estensen
Morton P. Friedman
John Gaito
W. H. Gladstones
Edward F. Gocka
Bert A. Goldman
William L. Hays
A. Hazewinkel
John K. Hemphill
Kenneth D. Hopkins
Edwin B. Hutchins
Paul I. Jacobs
J. A. Keats
Eric Klinger
Robert R. Knapp
S. David Leonard
G. A. Lienert
Jefferson F. Lindsey, Jr.
John Marshall Long
Milton H. Maier
James N. McClelland
Jason Millman
Richard Millward
Donald F. Morrison
Jane Srygley Mouton
Bishwa Nath Mukherjee
Thomas F. Nichols

Kazuo Nihira
Melvin R. Novick
Tapio Nummenmaa
LeRoy A. Olson
Robert Travis Osborne
Treadway C. Parker
James L. Pate
Jose F. Pisani
Eugene Richards
Raymond R. Ritti
Donald Clare Ross
John Ross
William H. Sammons
Johann M. Schepers
Elizabeth F. Shipley
Hirsch L. Silverman
Douglas Sjogren
Marvin Snider
Doris V. Springer
Saul H. Sternberg
Harry E. Stine
Peter H. Ten Eyck
Donald John Veldman
Leonard Wevrick
S. Wiegersma
Peter Wolmut
Kenneth Russell Wood
Henry J. Zagorski

It was moved, seconded, and passed that the 22 persons named below be elected as student members.

Robert L. Collins
Anna B. Cox
James N. Cronholm
Richard K. Eyman
Ram K. Gupta
John Leonard Horn
George P. Huff
Earl Jennings
Robert L. Jones
Emily B. Kirby
Kenneth R. Laughery
Edmond Marks
Suchoon S. Mo
Panna Lal Pradhan
Ralph L. Rosnow
Richard M. Singleton
Lennart Sjoberg
Per Sjostrand
S. Paul Slovic
Harry L. Snyder
Elizabeth T. Wooldridge
Selwyn A. Zerof

It was moved, seconded, and passed that the Membership Committee be thanked for their excellent work.

Dr. John E. Milholland reported for the Program Committee that 3 symposia and 4 paper reading sessions had been scheduled jointly with Division 5, in addition to the presidential address and social hour. It was moved, seconded, and passed that the report be accepted with thanks.

Dr. William B. Schrader reported for the Committee on Relations Between the Psychometric Society and the Psychometric Corporation as follows:

"Professor Irving Lorge, as Chairman of the Committee on Relations Between the Psychometric Society and the Psychometric Corporation, had virtually completed the preparation of the proposed new Certificate of Incorporation and By-Laws of the Psychometric Society at the time of his death in January, 1961.

"The Committee plans to take the following actions in preparation for the incorporation of the Society. First, submit the proposed new Constitution to all regular members in good standing by mail for a vote on acceptance or rejection. The mailing will be made on or about October 1, 1961. Second, report the results of the voting to the President of the Society. If two-thirds or more of the ballots received favor the new

Constitution, the report would outline the further steps needed to bring about the incorporation of the Society under the laws of New Jersey."

On motion, the report of this Committee was accepted with thanks.

Dr. William B. Schrader reported for the Auditing Committee. He stated that all financial matters of the Society were found to be in good order. The report was accepted with thanks.

The report of the Treasurer was presented by Dr. John W. French. A copy is attached. The report was accepted with thanks.

It was announced that Dr. Philip H. DuBois had been elected President of the Society for the term ending 30 September 1962.

On a ballot for the election of two new members of the Council of Directors, Dr. Warren S. Torgerson and Dr. Charles F. Wrigley were elected for a term of three years, ending in 1964.

On a ballot for the election of a secretary, Dr. William G. Mollenkopf was elected for a term of three years, ending 30 September 1964. He succeeds Philip H. DuBois whose resignation becomes effective 30 September 1961.

A resolution from the Council of Directors relative to the formation of a society with somewhat similar objectives as those of the Psychometric Society was discussed at length. On motion the resolution was tabled.

The meeting was adjourned at 10:50 a.m.

Philip H. DuBois
Secretary

PSYCHOMETRIC SOCIETY

Statement of Receipts and Disbursements for Fiscal Year Ended June 30, 1961

RECEIPTS

Dues:	Year	Members	Student Members	
	1962	1		
	1961	679	50	
	1960	70	24	
	1959	2		
		752	74	$5,560.00
Net partial payment of dues				3.60
			Total dues:	$5,563.60

Contribution to 25th Anniversary Fund	$ 1.00
Received with dues for Corporation Publications	156.60
Refund from Corporation for an overpayment	8.25
	$5,729.45

DISBURSEMENTS

Psychometric Corporation (90% of dues)	$5,007.24
Psychometric Corporation for publication collection	156.60
Stationery and postage	165.25
Secretarial services	172.67
Mailing services by Byrd Press	101.80
25th Anniversary Fund expenses	897.14
Refund	7.00
Lawyer's fee in connection with incorporation (50%)	37.50
Miscellaneous	13.29
Total disbursements	$6,558.49

BALANCE

Balance, June 30, 1960	$2,092.56
Receipts, 1960–61	5,729.45
	$7,822.01
Disbursements, 1960–61	6,558.49
Balance, June 30, 1961	$1,263.52

August 28, 1961

John W. French, Treasurer

PSYCHOMETRIC CORPORATION

Statement of Receipts and Disbursements for Fiscal Year Ended June 30, 1961

RECEIPTS	
Subscriptions (less agency discounts)	$ 8,195.60
Psychometric Society (90% of dues)	5,007.24
Sales of back issues (less discounts)	4,842.12
Sale of Monographs 5–8	123.60
Interest on Savings Accounts	681.72
Reprints	838.00
Fee for use of Society mailing list	10.00
	$19,698.28
DISBURSEMENTS	
Printing and Mailing *Psychometrika* Volume 25, No. 2, through 26, No. 1	$ 8,442.26
Reprints	686.60
Stipend of Managing Editor	1,500.00
Stipend of Assistant Editor	750.00
Stipend of Treasurer	500.00
Secretarial services: Editorial office	800.00
Secretarial services: Business office	104.60
Stationery and postage	258.44
Mailing back issues and monographs	219.23
Refunds	38.10
Reprinting Volume 2, No. 3	116.00
Lawyer's fee in connection with incorporation (50%)	37.50
Bond for treasurer (2 years)	50.00
Miscellaneous	82.25
	$13,584.98
BALANCE	
Balance, June 30, 1960	$ 8,249.82
Reserve Funds, June 30, 1960	
Englewood Savings and Loan	3,500.00
Metropolitan Savings and Loan	8,500.00
Total assets, June 30, 1960	$20,249.82
Receipts, 1960–61 (add)	19,698.28
	$39,948.10
Disbursements, 1960–61 (subtract)	13,584.98
Total assets, June 30, 1961	$26,363.12
DISPOSAL OF ASSETS	
Cash balance, June 30, 1961	$ 2,212.67
Reserve Funds, June 30, 1961	
Englewood Savings and Loan	3,500.00
Metropolitan Savings and Loan	8,500.00
First National Bank, Princeton	12,150.45
Total assets	$26,363.12
OBLIGATIONS	
Estimated cost of *Psychometrika* to end of year	$ 6,300.00
Stipends	1,375.00
Secretarial services and postage	550.00
	$ 8,225.00
ASSETS LESS OBLIGATIONS	$18,138.12

August 28, 1961 — John W. French, Treasurer

INDEX FOR VOLUME 26

P

B

VOL
NUM
DEC
9

Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY - ORGANIZED IN 1935

VOLUME 26
NUMBER 4
DECEMBER
1961

Psychometrika, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

DECEMBER, 1961, VOLUME 26, NUMBER 4

Published by the Psychometric Society at 1407 Sherwood Avenue, Richmond 5, Virginia. Second-class postage paid at Richmond, Virginia. Editorial Office, Department of Psychology, Purdue University, Lafayette, Indiana.

Subscription Price: The regular subscription rate is $14.00 per volume. The subscriber receives each issue as it comes out, and, upon request, a second complete set for binding at the end of the year. All annual subscriptions start with the March issue and cover the calendar year. Those back issues which remain available are $14.00 per volume (one set only) or $3.50 per issue, with a 20 percent discount to Psychometric Society members. Members of the Psychometric Society pay annual dues of $7.00, of which $6.30 is in payment of a subscription to *Psychometrika*. Student members of the Psychometric Society pay annual dues of $4.00, of which $3.60 is in payment for the journal.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to

ROBERT L. EBEL
Educational Testing Service
P. O. Box 586
Princeton, New Jersey

Payments: All bills and orders are payable in advance.

Checks covering membership dues should be made payable to the *Psychometric Society*.

Checks covering regular subscriptions to *Psychometrika* (for non-members of the Psychometric Society) and back issue orders should be made payable to the *Psychometric Corporation*. All checks, notices of change of address, and business communications should be sent to

JOHN W. FRENCH, Treasurer, Psychometric Society and Psychometric Corporation
Educational Testing Service
P.O. Box 586
Princeton, New Jersey

Articles on the following subjects are published in *Psychometrika:*

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as nomographs, tables, worksheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

The emphasis is to be placed on articles of type (1), insofar as articles of this type are available.

(*Continued on the back inside cover page*)

In the selection of the articles to be printed in *Psychometrika*, an effort is made to obtain objectivity of choice. All manuscripts are received by one person, who first removes from each article the name of contributor and institution. The article is then sent to three or more persons who make independent judgments upon the suitability of the article submitted. This procedure seems to offer a possible mechanism for making judicious and fair selections.

Prospective authors are referred to the "Rules for Preparation of Manuscripts for *Psychometrika*," contained in the March, 1958 issue. Reprints of these "Rules" are available from the managing editor upon request. A manuscript which fails to comply with these requirements will be returned to the author for revision.

Authors will receive 100 reprints without covers, free of charge.

Manuscripts for publication in *Psychometrika* should be sent to
B. J. Winer, Managing Editor, *Psychometrika*
Dept. of Psychology, Purdue University
Lafayette, Indiana

Material for review in *Psychometrika* should be sent to
John E. Milholland, Review Editor, *Psychometrika*
122 Rackham Bldg., Univ. of Michigan
Ann Arbor, Michigan

The officers of the Psychometric Society for the year October 1961 through September 1962 are as follows: *President:* Philip H. DuBois, Dept. of Psychology, Washington Univ., St. Louis 30, Missouri; *Secretary:* William G. Mollenkopf, 231 Hillcrest Ave., Cincinnati 15, Ohio; *Treasurer:* John W. French, Educational Testing Service, P. O. Box 586, Princeton, New Jersey.

The Council members, together with dates at which terms expire, are as follows: Jane Loevinger, 1962; John E. Milholland, 1962; Allen L. Edwards, 1963; Bert F. Green, 1963; Warren S. Torgerson, 1964; Charles F. Wrigley, 1964.

The *Psychometric Monographs Committee* is composed of Frederick B. Davis, *Chairman;* Harold Gulliksen, Paul Horst, and Frederic Kuder. Manuscripts and correspondence for this series should be addressed to

Frederick B. Davis, Chairman, Psychometric Monographs Committee,
Hunter College, 695 Park Avenue, New York 21, N. Y.

PRINTED BY THE WILLIAM BYRD PRESS, INCORPORATED, RICHMOND, VA.

2524